

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2002, 07:38:37 ; Search time 133 Seconds  
(without alignments)  
5514.763 Million cell updates/sec

Title: US-09-845-721-1  
Perfect score: 1881  
Sequence: 1 gttatttcttcaaaaggaaa.....aaaaaaaaaaaaaaaaaaaa 1881

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 350425 seqs, 194966369 residues

Total number of hits satisfying chosen parameters: 700850

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

#### Database :

Published\_Applications\_NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	72	3.8	416	10	US-09-960-352-4584
C 2	66.2	3.5	419	10	US-09-960-352-11234
C 3	59.6	3.2	406	10	US-09-960-352-10265
C 4	58.4	3.1	393	10	US-09-960-352-4582
C 5	58.4	3.1	2000	9	US-09-938-842A-2868
C 6	57.4	3.1	337	10	US-09-960-352-6976
C 7	57	3.0	377	10	US-09-960-352-7419
C 8	56.2	3.0	375	10	US-09-960-352-15014
C 9	54.4	2.9	277	10	US-09-960-352-12673
C 10	54.2	2.9	424	10	US-09-960-352-11218
C 11	53.4	2.8	241	10	US-09-960-352-5777
C 12	52.8	2.8	298	10	US-09-960-352-1004
C 13	52.8	2.8	302250	10	US-09-962-832-154
C 14	52.4	2.8	277	10	US-09-960-352-12673
C 15	52.4	2.8	419	10	US-09-960-352-11234
C 16	52	2.8	5332	10	US-09-801-861-3
C 17	51.8	2.8	375	10	US-09-960-352-15014
C 18	51.2	2.7	416	10	US-09-960-352-4584
C 19	50.4	2.7	4877	10	US-09-834-975-872

C 20	50.4	2.7	4877	10	US-09-834-975-873	Sequence 873, App
C 21	50.4	2.7	4877	10	US-09-834-975-884	Sequence 884, App
C 22	50.2	2.7	380	10	US-09-960-352-9335	Sequence 9335, App
C 23	50.2	2.7	1797	10	US-09-729-674-89	Sequence 89, Appl
C 24	50	2.7	408	10	US-09-960-352-6263	Sequence 6263, App
C 25	49.8	2.6	397	10	US-09-960-352-13784	Sequence 13784, A
C 26	49.6	2.6	1371	10	US-09-764-853-227	Sequence 227, App
C 27	49.4	2.6	358	10	US-09-960-352-5559	Sequence 5559, App
C 28	49.4	2.6	390	10	US-09-960-352-1976	Sequence 1976, App
C 29	49.4	2.6	640681	10	US-09-790-988-1	Sequence 1, Appl
C 30	49.2	2.6	932	12	US-10-078-929-93	Sequence 93, Appl
C 31	49.2	2.6	1377	10	US-09-925-299-93	Sequence 2, Appl
C 32	49.2	2.6	1931	10	US-09-748-537-2	Sequence 2, Appl
C 33	49.2	2.6	1931	10	US-09-728-721-1	Sequence 1, Appl
C 34	49.2	2.6	1931	12	US-10-133-780-2	Sequence 2, Appl
C 35	49.2	2.6	1931	12	US-10-105-931-1	Sequence 1, Appl
C 36	49	2.6	241	10	US-09-960-352-7904	Sequence 7904, App
C 37	49	2.6	368	10	US-09-834-975-47	Sequence 47, Appl
C 38	49	2.6	425	10	US-09-834-975-451	Sequence 451, App
C 39	49	2.6	2645	10	US-09-764-870-76	Sequence 76, Appl
C 40	48.8	2.6	2000	9	US-09-938-842A-3939	Sequence 3939, App
C 41	48.6	2.6	376	10	US-09-960-352-5087	Sequence 5087, App
C 42	48.6	2.6	302250	10	US-09-962-832-154	Sequence 154, App
C 43	48.4	2.6	415	10	US-09-969-373-751	Sequence 751, App
C 44	48	2.6	2890	10	US-09-925-300-763	Sequence 763, App
C 45	48	2.6	180557	12	US-10-003-806-6	Sequence 6, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-960-352-4584/c

; Sequence 4584, Application US/09960352

; Patent No. US20020137139A1

; GENERAL INFORMATION:

; APPLICANT: Warren, Wesley C.

; APPLICANT: Byatt, John C.

; APPLICANT: Mathalagan, Nagappan

; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION

; FILE REFERENCE: 16511.006/37-21(10298)C

; CURRENT APPLICATION NUMBER: US/09/960,352

; CURRENT FILING DATE: 2001-09-24

; NUMBER OF SEQ ID NOS: 15112

; SEQ ID NO 4584

; LENGTH: 416

; TYPE: DNA

; ORGANISM: Bos taurus

; OTHER INFORMATION: Clone ID: 20-LIB3057-016-Q1-K1-E11

US-09-960-352-4584

Query Match

Best Local Similarity 3.8%; Score 72; DB 10; Length 416;

Matches 201; Conservative 0; Mismatches 215; Indels 0; Gaps 0;

QY 1466 AACACAAAAAATTACCTTTGAAGATAGTGAATAAGAGAAAAACGTTTGTAGTCCCTCAGGT 1525

Db 416 AAAAAAATTTTTTTTTTAAATAAATTTTTTATATAAATTTTTTAAATTTATAAAT 357

QY 1526 TGTACAGACTAGAGAAAGCTCAGTTTCACCAATCCACATCAATGAGTTTAAAT 1585

Db 356 TACCATATATTTTTTAAATATTTTTTAAATAAATCTTTTTTATATCTCTATAAT 297

QY 1586 TTAATTTGTAATAACTGATATTACTGCCAAATATAAGAAAAATATTTTAAGCTATTGGTTA 1645

Db 296 TAATAAATTTTAAATTTTAAATAAATTTTCTTTAAATAATTTTAAATTTTAAAA 237

QY 1646 TGTGTAATTTTCAATGTGAATGCTAAATAGATAGCTATATATATCAATTTCTTCA 1705

Db 236 TATTTTTTTTTTGAATTTTTTAAATTTTATATATTTTTTAAATTTTAAATCTTTTT 177





```

; SEQ ID NO 7419
; LENGTH: 377
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION:
US-09-960-352-74

```

Query Match 3.0%; Score 57; DB 10; Length 377;  
Best Local Similarity 52.2%; Pred. No. 0.096;  
Matches 155; Conservative 0; Mismatches 135; Indels 7; Gaps 1;

Qy	1585	TTTTAAATTTGATAAAACTGATATTTACTGCCAAATAATAGACAAAAATATTTTAAATATTCGGTT	1644
Db	317	TTTATGTTGTTGGACCGTTTTTTTTTTTTTTTTTTTGAAGAATTTTTTTAGTGTATTTTTTTTTTGTATAT	258
Qy	1645	ATGTTCTAAATTTTTCAAATGTGAATGCTAAATAGATAGGTGCATATATATCAAAATTCCTTC	1704
Db	257	ATGCTTTATATTTATATATTTTATATTTTATATTTGATTTAGTATTTTATTTTAAATTTAATGT	198
Qy	1705	ATTACCTTAATGTATTTGTTGCAATGGCAGTTTGTTAAAGTACTATCATCTGTATATATTTGTT	1764
Db	197	TTGTTTTTAATATTTTATTTAAATGTTGGATTTTTTTTATTTTCATTAATTTTAAATTTTTTT	138
Qy	1765	CAATATTTATGTCCACAGAAAAATTTTCATGTAAAGTCATATTTTTTAAAGGAATAAATACAT	1824
Db	137	ATTAAATTTATATATTTTTTTTTTAAATTAATTAAT	85
Qy	1825	AGCCTTAAACAGCTGTATTAACCTTTTAAATGTAAAAAATAAAAAAAAAAAAAAAAAAAAAA	1881
Db	84	AAAAATATAAAAATATATAAAAAACAACCAAAAAAATAAAAAAAAAAAAAAAAAAAAAA	28

```

RESULT 8
US-09-960-352-15014/c
; Sequence 15014, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 15014
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 64-LIB3058-048-Q1-K1-H8
US-09-960-352-15014

```

Query Match 3.0%; Score 56.2; DB 10; Length 375;  
Best Local Similarity 51.8%; Pred. No. 0.13;  
Matches 127; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

[illegible]

QY 1877 AAAAA 1881  
Dbb 70 AAAAA 66

```

RESULT 9
US-09-960-352-12673/c
; Sequence 12673, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalaagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCES: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12673
; LENGTH: 277
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 54-LIB3058-039-Q1-K1-F10
US-09-960-352-12673

```

```
Query Match      2.9%; Score 54.4; DB 10; Length 277;
Best Local Similarity 52.7%;
Matches 118; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
```

[illegible]

```

RESULT 10
US-09-960-352-11218/c
; Sequence 11218, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11218
; LENGTH: 424
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-LIB3058-026-Q1-K1-D12
US-09-960-352-11218

```

Query Match 2.9%; Score 54.2; DB 10; Length 424;  
Best Local Similarity 50.5%; Pred. No. 0.3;

Matches 161; Conservative 0; Mismatches 153; Indels 5; Gaps 1;

QY 1568 TTCAAATGAGCTTTAAATTTAAATTTGTAAGAACTGATATTACTGCCAAATATAGAAGAAA 1627  
Db 354 TTTTATATCTCTTTATTTTAAATTTTATATCTTTTTTTTTTAAATAACTTAATTC 295  
QY 1628 TATTTTAAGTATGGTATCTGTAAATTTTCAATGGAATGCTAA-----TTAGATAG 1682  
Db 294 TTTTAAAAATTTCAATTTTTTTTTTTTCTTTAAATATCTCTTTTATATTTTTTTTAAAT 235  
QY 1683 GTCATATATATCAATTTCTTCATTAATGATATTTGTCATGCGACGTTTGTAAAG 1742  
Db 234 TCCTTATAATATTTTATATATTTTAAACAAATATTTAAATTTTAAATTTATTTAT 175  
QY 1743 TACTATCATGTGATATTTTGTCAATATATGTCACACAGAAAATATTCATGTAAGTCAT 1802  
Db 174 TTTTAAATTTTATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 115  
QY 1803 ATTTTAAAGGAATAATACATAGCCTTAAACAGTGATATACTTTAAATGTAAAGAAA 1862  
Db 114 TTTTATTTTAAATTTAAAGAAAACCAACCAACCAACCAACCAACCAACCAACCA 55  
QY 1863 AAAAAAAAAAAAAAAAAAAAAA 1881  
Db 54 AAAAAAAAAAAAAAAAAAAAAA 36

RESULT 11  
US-09-960-352-5777/c  
; Sequence 5777, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; PRIOR FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 5777  
; LENGTH: 241  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 25-LIB3057-009-Q1-K1-G1  
US-09-960-352-5777

Query Match 2.8%; Score 53.4; DB 10; Length 241;  
Best Local Similarity 69.9%; Pred. No. 0.37; Indels 0; Gaps 0;  
Matches 72; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1779 ACAGAAATATTCATGTAAGTCATATTTTAAAGGAATAATACATAGCCTTAAACAGT 1838  
Db 233 ATAAATAATTTAAATAAATTTAAATTTTAAATAATATTTTAAATAAATAATATA 174  
QY 1839 GTATACTTTTAAATGTAAAGAAAACCAACCAACCAACCAACCAACCAACCAACCA 1881  
Db 173 TTATAAAATTAATTTAAAGAAAACCAACCAACCAACCAACCAACCAACCAACCAACCA 131

RESULT 12  
US-09-960-352-1004/c  
; Sequence 1004, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; PRIOR FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 1004  
; LENGTH: 298  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 05-LIB3057-009-Q1-K1-B1  
US-09-960-352-1004

Query Match 2.8%; Score 52.8; DB 10; Length 298;  
Best Local Similarity 56.2%; Pred. No. 0.49; Indels 0; Gaps 0;  
Matches 99; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1706 TTACTTAATGATATTTGTCATGCGACGTTTGTAAAGTACTATCATGTCATATTTTGTGTC 1765  
Db 218 TTTCTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 159  
QY 1766 AATATATGTCACACAGAAAATATTCATGTAAGTCATATTTTAAAGAAATAAATACATA 1825  
Db 158 TTTTATTTTAAATAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 99  
QY 1826 GCCTTAAACAGTGATATACTTTTAAATGTAAAGAAAACCAACCAACCAACCAACCAACCA 1881  
Db 98 AAAAAAAAAAAAAAAAAAAAAA 43

RESULT 13  
US-09-962-832-154/c  
; Sequence 154, Application US/09962832  
; Patent No. US20020110821A1  
; GENERAL INFORMATION:  
; APPLICANT: Ebner, Reinhard  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign  
; FILE REFERENCE: 689290-74  
; CURRENT APPLICATION NUMBER: US/09/962,832  
; CURRENT FILING DATE: 2001-09-25  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,077  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,280  
; NUMBER OF SEQ ID NOS: 259  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 154  
; LENGTH: 302250  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-962-832-154

Query Match 2.8%; Score 52.8; DB 10; Length 302250;  
Best Local Similarity 52.1%; Pred. No. 2.2;  
Matches 165; Conservative 0; Mismatches 149; Indels 3; Gaps 2;

QY 1567 ATCAATGAGTTTTTAAATTTAAATTTGTAAGAACTGATATTACTGCCAAATATAGAAGAAA 1626  
Db 300411 ATATATAATCTCTTATATATATTTTATATAATTTTAAATAATAATAATAATATATAT 300352  
QY 1627 ATATTTTAACTGTTGTTGTAATTTTCAATGTAAGTAACTGTAAGTAACTGTAAGTAACTGTA 1686  
Db 300351 ATATAATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTATATA 300292  
QY 1687 TATATATTCATTTCTTCAT--TACTTAATGTTTGTGTCATGCGACGTTTGTCTTAAAGTAC 1745  
Db 300291 TATATAATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTATATA 300232  
QY 1746 TATCATGTGATATTTTGTCA--ATATATGTCACACAGAAAATATTCATGTAAGTACATA 1803  
Db 300231 TATTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATATTTTATATA 300172  
QY 1804 TTTTAAAGAAATAAATACATAGCCTTAAACAGTGTAATACTTTAAAGTAAAGTAAAGTAAAGTAA 1863



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2002, 07:35:27 ; Search time 2401 seconds

(without alignments)

12687.926 Million cell updates/sec

Title: US-09-845-721-1

Perfect score: 1881

Sequence: 1 gttatttcttcaaaaggaaa.....aaaaaaaaaaaaaaaaaaaaa 1881

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: em_estha:*
2: em_esthum:*
3: em_estinu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_esttc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	609.4	32.4	684	17 B56597	B56597 CIT-HSP-200
2	576.2	30.6	761	12 BE824968	BE824968 C0507E09-
3	552.6	29.4	938	12 BF968206	BF968206 602268947
4	541.2	28.8	984	12 BG295999	BG295999 602395255
5	497.8	26.5	668	12 BE824971	BE824971 C0507E12-
6	474	25.2	631	12 BE824600	BE824600 C0501H07-

7	471	25.0	647	10	BB171739	BB171739
8	462.6	24.6	570	17	AZ494191	AZ494191 IM0329L18
9	451.8	24.0	519	12	BG086308	BG086308 H3124D01-
10	435.8	23.1	587	12	BE824614	BE824614 C0502803-
11	435	23.1	591	12	BE824967	BE824967 C0507E08-
12	431.6	22.9	587	12	BE824766	BE824766 C0504E03-
13	398.4	21.2	559	12	BF075992	BF075992 225541 MA
14	379	20.1	422	9	AL600185	AL600185 DRFP313J
15	362.6	19.3	520	10	AW061316	AW061316 UT-M-BH1-
16	346.4	18.4	488	12	BE824969	BE824969 C0507E10-
17	340.6	18.1	465	17	AQ222591	AQ222591 HS_2007_B
18	326	17.3	345	14	Z18870	Z18870 HSDHEG001 S
19	320.6	17.0	463	9	AA500806	AA500806 v901e07.r
20	319	17.0	459	12	BE824472	BE824472 C0500A07-
21	313.6	16.7	460	12	BE824814	BE824814 C0505C01-
22	296.8	15.8	461	14	R59799	R59799 Yh07h07.r1
23	289.6	15.4	509	14	R61341	R61341 Yh07h07.s1
24	244.4	13.0	258	14	R58357	R58357 G3141 Retal
25	240.8	12.8	682	10	BB366337	BB366337
26	223.6	11.9	568	17	AQ608397	AQ608397 HS_2122_B
27	213.6	11.4	333	9	AV147896	AV147896 AV147896
28	184	9.8	266	9	AI614903	AI614903 v901e07.y
29	180.4	9.6	304	10	AW122633	AW122633 UT-M-BH2
30	177	9.4	284	10	BB310100	BB310100 BB310100
31	172.2	9.2	348	10	BB259564	BB259564 BB259564
32	167.8	8.9	330	10	BB131127	BB131127 BB131127
33	167.4	8.9	320	10	BB464297	BB464297 BB464297
34	162.6	8.6	281	14	T20318	T20318 EST009 ED18
35	161.6	8.6	261	10	AV329383	AV329383 AV329383
36	161	8.6	337	10	BB314194	BB314194 BB314194
37	152.6	8.1	235	10	BB176012	BB176012 BB176012
38	149.6	8.0	1012	17	CNS02MDL	AL203970 Tetraodon
39	147	7.8	261	10	BB191298	BB191298 BB191298
40	146.6	7.8	717	9	AI641710	AI641710 fc22h07.x
41	144.6	7.7	235	10	AV341872	AV341872 AV341872
42	142.6	7.6	387	17	B38040	B38040 HS-1048-B2-
43	139.2	7.4	239	10	BB269323	BB269323 BB269323
44	137.6	7.3	290	10	BB188782	BB188782 BB188782
45	134.2	7.1	250	10	BB431762	BB431762 BB431762

#### ALIGNMENTS

RESULT 1	B56597	B56597	684 bp	DNA	linear	GSS 20-JUN-1998
LOCUS	CIT-HSP-2007F6.TFC	CIT-HSP Homo sapiens	genomic clone	2007F6	DNA	
DEFINITION	sequence.					
ACCESSION	B56597					
VERSION	B56597.1	GI:2610931				
KEYWORDS	GSS.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1 (bases 1 to 684)					
AUTHORS	Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.					
TITLE	Use of a random BAC End Sequence Database for Sequence-Ready Map Building					
JOURNAL	Unpublished (1997)					
COMMENT	Other GSSs: CIT-HSP-2007F6.TR					
	Contact: Mark Adams					
	Department of Eukaryotic Genomics					
	The Institute for Genomic Research					
	9712 Medical Center Dr., Rockville, MD 20850, USA					
	Tel: 301 838 0200					
	Fax: 301 838 0208					
	Email: mdadams@igir.org					
	Clones are available from Research Genetics (info@resgen.com). BAC					
	end search page:					

http://www.tigr.org/tdb/hungen/bac\_end\_search/bac\_end\_search.html  
Seq primer: M13-21  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1. .684  
/organism="Homo sapiens"  
/db\_xref="CDB:7040000"  
/db\_xref="taxon:9606"  
/clone="2007F6"  
/clone\_lib="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII"

BASE COUNT 222 a 129 c 94 g 239 t

ORIGIN

Query Match 32.4%; Score 609.4; DB 17; Length 684;  
Best Local Similarity 99.8%; Pred. No. 3e-74; Indels 0; Gaps 0;  
Matches 610; Conservative 0; Mismatches 1;

Qy 1 GTTATTTCTTCAAGAGAAACACAAATTTCTTTATATCAAAACAATGCAAACTTGATG 60  
Db 74 GTTATTTCTTCAAGAGAAACACAAATTTCTTTATATCAAAACAATGCAAACTTGATG 133  
Qy 61 GTTCTTAATTTCTACATTTCTTATTAATAGTTTACAACTTAAATTAAGTACAC 120  
Db 134 GTTCTTAATTTCTACATTTCTTATTAATAGTTTACAACTTAAATTAAGTACAC 193  
Qy 121 AATTGAAGATTTTCTTACAAACACACGTTATACGTCATTTAAATGGCAATA 180  
Db 194 AATTGAAGATTTTCTTACAAACACACGTTATACGTCATTTAAATGGCAATA 253  
Qy 181 TCAATAGTTTATTTTATTTTCTTCTAGGAAAAAACCACTGCTCCAAAAGAATGT 240  
Db 254 TCAATAGTTTATTTTATTTTCTTCTAGGAAAAAACCACTGCTCCAAAAGAATGT 313  
Qy 241 GTTTTCTCCGATTTCTGGAATCAACATGAGTCGTAATCTAACATTCACAGTGGAGATG 300  
Db 314 GTTTTCTCCGATTTCTGGAATCAACATGAGTCGTAATCTAACATTCACAGTGGAGATG 373  
Qy 301 ACATTGATGACATCAACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTCAG 360  
Db 374 ACATTGATGACATCAACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTCAG 433  
Qy 361 TGTCTCTCACCAGGATTTCTTATGTAGAAATTTGTTGGGACTTGGCAGCAACCTCACTG 420  
Db 434 TGTCTCTCACCAGGATTTCTTATGTAGAAATTTGTTGGGACTTGGCAGCAACCTCACTG 493  
Qy 421 TATTGGTACTTTACTGATGAATCCAACTTAATCAACTCTGTCAGTAAACATTAATACAA 480  
Db 494 TATTGGTACTTTACTGATGAATCCAACTTAATCAACTCTGTCAGTAAACATTAATACAA 553  
Qy 481 TGAATCTTCATGTTACTGATGAATTAATTTCTGTTGGGATGATTCCTCTAAGTATAGTTA 540  
Db 554 TGAATCTTCATGTTACTGATGAATTAATTTCTGTTGGGATGATTCCTCTAAGTATAGTTA 613  
Qy 541 TCCCTCTCTTCTACTGGAGAGTAACTGCTCTCATTTGCTGTTTCCATGAGGCTTGTG 600  
Db 614 TCCCTCTCTTCTACTGGAGAGTAACTGCTCTCATTTGCTGTTTCCATGAGGCTTGTG 673  
Qy 601 TATCTTTTGTCA 611  
Db 674 TATCTTTTGTCA 684

RESULT 2  
BE824968/c 761 bp mRNA linear EST 21-SEP-2000  
LOCUS BE824968  
DEFINITION C0507E09-3 NIA Mouse E13.5 VMB Dopamine cell cDNA Library Mus  
musculus cDNA clone C0507E09 3', mRNA sequence.  
ACCESSION BE824968  
VERSION BE824968.1 GI:10257202

KEYWORDS  
SOURCE

house mouse.  
Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 761)  
Barrett,T., Xie,T., Piao,Y., Dillon-Carter,O., Kargul,G.J., Lim  
M.K., Chrest,F.J., Wersto,R., Rowley,D.L., Juhaszova,M., Zhou,L.,  
Vawter,M.P., Becker,K.G., Cheadle,C., Wood,W.H. III, McCann,U.D.,  
Freed,W.J., Ko,M.S.H., Ricaurte,G. and Donovan,D.M.  
Genes expressed by midbrain dopamine neurons: a dopamine  
neuron-specific cDNA library and microarray used to examine changes  
in gene expression during development and methamphetamine  
neurotoxicity

TITLE

Unpublished (2000)  
Contact: George J. Kargul  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6920, USA  
Email: cdna@lgsun.grc.nia.nih.gov  
niaEST (http://lgsun.grc.nia.nih.gov/cDNA/cDNA.html)

JOURNAL

COMMENT

COMMENT

Seq primer: -21M13 Forward  
High quality sequence stop: 761  
POLYA=Yes.

FEATURES  
source

Location/Qualifiers  
1. .761  
/organism="Mus musculus"  
/strain="TH-beta-gal transgenic mouse"  
/db\_xref="niaEST:C0507E09-3"  
/db\_xref="taxon:10090"  
/clone="C0507E09"  
/clone\_lib="NIA Mouse E13.5 VMB Dopamine cell cDNA  
Library"  
/sex="Male/Female"  
/dev\_stage="13.5dpc"  
/lab\_host="DH10B"

/note="vector: pSPORT1 (Gibco/BRL Life Technology);  
Site\_1: SalI; Site\_2: NotI; Total RNAs were extracted from  
3000 Dopamine cells (cell collected by Dr.Tanya Barrett).  
The double-stranded cDNA was synthesized by Gibco's kit  
with an Oligo(dT) primer [NotI primer-adaptor from  
Gibco/BRL]  
[5'-pgAGTAGTTCTAGATCGGCGGCCCTTTTTTTTTTTT-3'] from  
0.91ug of total RNA. The double-stranded cDNAs were  
treated with T4 DNA polymerase and purified by  
ethanol-precipitation. The cDNAs were ligated to  
Lone-linker LL-Sal4 (include SalI sequence). The cDNAs  
were purified by phenol/chloroform and separated from free  
linkers by Centricon 100. Then, cDNAs were amplified by  
long-range high fidelity PCR using Takara's Ex Taq  
polymerase. Then, the cDNAs were purified by  
phenol/chloroform and by Centricon 100. The cDNAs were  
digested with SalI and NotI enzymes. Then, the cDNAs were  
size selected by Gibco's Size Fractionation Column. The  
cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid  
vector. The DH10B E. coli host was transformed with the  
ligation mixture by chemical method. The library was  
constructed by Yulan Piao and Minoru.S. H. Ko."

BASE COUNT

244 a 139 c 128 g 250 t

ORIGIN

Query Match 30.6%; Score 576.2; DB 12; Length 761;  
Best Local Similarity 87.4%; Pred. No. 9.3e-70;  
Matches 567; Conservative 0; Mismatches 88; Indels 8; Gaps 3;  
Qy 1074 GTAGTCTTTTGGTGTAAAGAACTTCAGTTCTGTGTAATAATTGCCCTCCGCGAGCTGTGAAA 1133  
Db 761 GTCTGTTTGGTGTGAGAACTTCAGTTCTCCGTAATAAATTCGCTCCGCGAGCGGTGAAG 702  
Qy 1134 CGACACCTGTGACGACGAG 1193  
Db 701 GCGCCACCGGACCGACGAG 642



```
QY 1194 TCTACATTTCTCTCTGCTGGACACCAATTTCTGTTTAAATACCACCATTTATGTTTA 1253
Db 1195 TCTACATTTCTCTCTGCTGGACACCAATTTCTGTTTAAATACCACCATTTATGTTTA 582
QY 1254 GGCCCAAGTGACCTTTTAGTAAATTAAGATTTGTTTATTTAGTCATGGCTTATGGAACA 1313
Db 1255 GGCCCAAGTGACCTTTTAGTAAATTAAGATTTGTTTATTTAGTCATGGCTTATGGAACA 522
QY 1314 ACTATATTTCCACCTCTATATATATATATATATATATATATATATATATATATATAT 1373
Db 1315 ACTATATTTCCACCTCTATATATATATATATATATATATATATATATATATATATAT 462
QY 1374 AGTAAATGAAGAAAGCGAGTTGTTTCTATAGTAGAAGCTGATCCCTGCCTCAATAAGCT 1433
Db 1375 AGTAAATGAAGAAAGCGAGTTGTTTCTATAGTAGAAGCTGATCCCTGCCTCAATAAGCT 402
QY 1434 GTAATACACAACTCTTGGATAGATCCCAAGAAACAAACAAATTTACCTTTGAAGATAGT 1493
Db 1435 GTAATACACAACTCTTGGATAGATCCCAAGAAACAAACAAATTTACCTTTGAAGATAGT 342
QY 1494 GAAATAGAGAAAACGTTTAGTGCCCTCAGGTTGTCACAGACTAGAGAAAAGTCTCAGTT 1553
Db 1495 GAAATAGAGAAAACGTTTAGTGCCCTCAGGTTGTCACAGACTAGAGAAAAGTCTCAGTT 282
QY 1554 TCACCAATCCACATTCGAATGAGTTTAAATTTAAATTTGTAAGAACTGATATTACTGCC 1613
Db 1555 TCACCAATCCACATTCGAATGAGTTTAAATTTAAATTTGTAAGAACTGATATTACTGCC 224
QY 1614 AAATATAGAAAATATTTTAAATTTGCTTATTTGTTAAATTTCAATTTGTAAGAACTGTA 1673
Db 1615 AAATATAGAAAATATTTTAAATTTGCTTATTTGTTAAATTTCAATTTGTAAGAACTGTA 164
QY 1674 ATTAGATAGGTCATATATATATATATATATATATATATATATATATATATATATAT 1729
Db 1675 ATTAGATAGGTCATATATATATATATATATATATATATATATATATATATATATAT 104
QY 1730 CAGTTTCTGTTTAAAGTAAATATATATATATATATATATATATATATATATATATAT 1787
Db 1731 CAGTTTCTGTTTAAAGTAAATATATATATATATATATATATATATATATATATATATAT 44
QY 1788 ATTATGTAAGTCATATATATATATATATATATATATATATATATATATATATATAT 1830
Db 1789 ATTATGTAAGTCATATATATATATATATATATATATATATATATATATATATATAT 1

RESULT 3
LOCUS BF968206 938 bp mRNA linear EST 22-JAN-2001
DEFINITION 602268947F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4357426 5',
mRNA sequence.
ACCESSION BF968206
VERSION BF968206.1 GI:12335421
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 938)
NIH-MGC http://mgi.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM993 row: m column: 11
High quality sequence stop: 711.
```

```
FEATURES
source
Location/Qualifiers
1..938
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4357426"
/lab_host="NIH_MGC_84"
/tissue_type="adrenal cortex carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site_1:
Noti; Site_2: SalI; Cloned unidirectionally; oligo-dT
primed. Average insert size 1.229 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 286 a 208 c 156 g 288 t
ORIGIN
Query Match 29.4%; Score 552.6; DB 12; Length 938;
Best Local Similarity 88.3%; Pred. No. 1.3e-66;
Matches 658; Conservative 0; Mismatches 79; Indels 8; Gaps 5;
QY 112 TAAGTACAAATTCGAAGATTTTTCCTTACAAAGAACACGTTATACGTCATTTAAAT 171
Db 113 TAAATCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 60
QY 172 TGCCAAATATCAATAGTTTATTTCTTCTAGGGAAGAAACCAACTGCTCCA 231
Db 173 CATGAATATAAACACAGCTACTTC-----ACTGTTGTCAGGAAAGAAACCAACTGCTCCA 116
QY 232 AAAGAATGTGTTTTCTCCCATTTCTGGAATCAACATGCACTGCAATCAACATTACAG 291
Db 117 AAAGAATGTGTTTTCTCCCATTTCTGGAATCAACATGCACTGCAATCAACATTACAG 176
QY 292 TGCAGATGACATTTGATGACATCAACACCAATATGTACCAACCACTATCATATCGTTAA 351
Db 177 TGCAGATGACATTTGATGACATCAACACCAATATGTACCAACCACTATCATATCGTTAA 236
QY 352 GCTTTCAAGTCTCTCTACCGGATTTCTTATGTTAGAAATTTGTTGGAGCTTGGCACA 411
Db 237 GCTTTCAAGTCTCTCTACCGGATTTCTTATGTTAGAAATTTGTTGGAGCTTGGCACA 296
QY 412 ACCTCACTGATTTGGTACTTTACTGCAATCAACCAATCAACCTTAATCAACTCTGTCAGTAA 471
Db 297 ACCTCACTGATTTGGTACTTTACTGCAATCAACCAATCAACCTTAATCAACTCTGTCAGTAA 356
QY 472 TTATTAATGAATCTTCATGCTACTTGTATGTAATTAATTTGTTGGGATGATTCCTCTAA 531
Db 357 TTATTAATGAATCTTCATGCTACTTGTATGTAATTAATTTGTTGGGATGATTCCTCTAA 416
QY 532 CTATAGTTATCTCTCTGCTTTCACTGGAGAGTAACACGCTCTCTCATTTTGTGTTTCCATG 591
Db 417 CTATAGTTATCTCTCTGCTTTCACTGGAGAGTAACACGCTCTCTCATTTTGTGTTTCCATG 476
QY 592 AGGCTTGTGATCTTTTGCAGAGTGTCTCAACAGCAATCAACGCTTTTGTCTATCACTTTGG 651
Db 477 AGGCTTGTGATCTTTTGCAGAGTGTCTCAACAGCAATCAACGCTTTTGTCTATCACTTTGG 536
QY 652 ACAGATGACATCTCTGTAACCACTGCAACCGCAATTTCTGACAAATGGGAGAGCTGT-A 710
Db 537 ACAGATGACATCTCTGTAACCACTGCAACCGCAATTTCTGACAAATGGGAGAGCTGTAA 596
QY 711 ATGTTAATGATATCCATTTGGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 769
Db 597 ATGTTAATGATATCCATTTGGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 656
QY 770 GGTAAATTTTTCAGTCTTCAAGTGGAAATACCTTGGGAAACCAACAGACACTTTTATGTGT 829
Db 657 GGTCAA-TTTTTCAGTCTTCAAGTGGAAATACCTTGGGAAACCAACAGACACTTTTATGTGT 714
QY 830 CAGTACAAATGAATTAATCACTGAA 854
Db 715 ACCATTGAATTAATTAATCACTGAA 739
```









QY 1175 GATGCTTTTATTGATTTTACATTTCTCTCTGCTGGACACCAATTTCTGTTTAA 1234  
|||||  
Db 153 AATGCTGTTATTGATTTTACATTTCTCTCTGCTGGACACCAATTTCTGTTTAA 94  
|||||  
QY 1235 TACCACCATTTTATGTTTAGGCCCAAGTGACCTTTTAGTAAATTAAGATGTTGTTTTT 1294  
|||||  
Db 93 TACCACCATTTCTGTTTAGGCCCAAGTGACCTTTTAGTAAATTAAGATGTTGTTTTCT 34  
|||||  
QY 1295 AGTCATGGCTTATGGAAACAATATATTTTCACCC 1327  
|||||  
Db 33 AGTCATGGCTTATGGAAACAATATATTTTCACCC 1  
|||||

RESULT 9  
LOCUS BG086308  
DEFINITION H3124D01-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone  
H3124D01 5', mRNA sequence.  
ACCESSION BG086308  
VERSION BG086308.1 GI:12568872  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jaradat, S.A., Tanaka  
J.S., Carter, M.G. and Ko, M.S.H.  
Verification and initial annotation of NIA mouse 15K cDNA clone set  
Other\_ESTs: H3124D01-3  
Unpublished (2001)  
Contact: George J. Kargul  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@igsun.grc.nia.nih.gov  
This clone set has been freely distributed to the community. Please  
visit <http://lgsun.grc.nia.nih.gov/cDNA/15k.html> for details.  
plate: H3124 row: D column: 01  
Seq primer: -21M13 Reverse  
High quality sequence stop: 519  
POLYA-No.

FEATURES  
source  
1..519  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="niaEST:H3124D01-5"  
/db\_xref="taxon:10090"  
/clone="H3124D01"  
/clone\_lib="NIA Mouse 15K cDNA Clone Set"  
/sex="Clones arrayed from a variety of cDNA libraries"  
/dev\_stage="Clones arrayed from a variety of cDNA  
libraries"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; This  
clone is among a rearranged set of 15,247 clones from 11  
embryo cDNA libraries (including preimplantation stage  
embryos from unfertilized egg to blastocyst, embryonic  
part of E7.5 embryos, extraembryonic part of E7.5 embryos  
, and E12.5 female mesonephros/gonad) and one newborn  
ovary cDNA library. Average insert size 1.5 kb. All  
source libraries are cloned unidirectionally with Oligo(dT  
)-Not primers. References include: (1) Genome-wide  
expression profiling of mid-gestation placenta and embryo  
using a 15,000 mouse developmental cDNA microarray, 2000,  
Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2)  
Large-scale cDNA analysis reveals phased gene expression  
patterns during preimplantation mouse development, 2000,  
Development, 127: 1737-1749; (3) Genome-wide mapping of  
unselected transcripts from extraembryonic tissue of  
7.5-day mouse embryos reveals enrichment in the t-complex  
and under-representation on the X chromosome, 1998, Hum

Mol Genet 7: 1967-1978.  
BASE COUNT 135 a 120 c 93 g 171 t  
ORIGIN  
Query Match 24.0%; Score 451.8; DB 12; Length 519;  
Best Local Similarity 91.9%; Pred. No. 1.1e-52;  
Matches 477; Conservative 0; Mismatches 42; Indels 0; Gaps 0;  
QY 288 ACAGTCGGAGATGACATTTGATGACATCAACACCAATATGTACCACCACTATCATATCG 347  
|||||  
Db 1 ACGGTGCGAGATGACATTTGATGACATCGACCAATATGTACCACCACTATCATATCG 60  
|||||  
QY 348 TTAAGCTTTCAAGTGTCTCTCACCGGATTTCTATCTTAGAAATTTCTGTTGGGACTTGC 407  
|||||  
Db 61 CTAAGCTTTCAAGTGTCTCTCTCACTGGATTTCTCATGTTAGAGATCGTCTGGGCGTTGC 120  
|||||  
QY 408 AGCAACCTCAGTATGTTGTTACTTACTGCATGAATCAACCTTAATCAACTCTGTCACT 467  
|||||  
Db 121 AGCAACCTTACCGTTCTGGTACTTTACTGCATGAATCAACCTTAATCAACTCTGTCACT 180  
|||||  
QY 468 AACATTATTACAATGAATCTTCATGTACTTACATGAATTAATTTGCTGGGATGATTCCT 527  
|||||  
Db 181 AACATTATTACAATGAATCTTCATGTACTTACATGAATTAATTTGCTGGGATGATTCCT 240  
|||||  
QY 528 CTAATATATAGTATCTCTCTGCTTTTCACTGGAGAGTAACACTGCTCTCATTTGCTGTTTC 587  
|||||  
Db 241 CTAATATAGTATCTCTCTGCTTTTCACTGGAGAGTAACACTGCTCTCATTTGCTGTTTC 300  
|||||  
QY 588 CATGAGCTTGTGTATCTTTTGCAGTGTCTCAACAGCAATCAACGTTTTTCTGTAFCAT 647  
|||||  
Db 301 CACGAAGCTTGTGTCTTCCCTTTGCAAGTGTTCGACAGCAATCAACGTTTTTCTGTAFCAT 360  
|||||  
QY 648 TTGGACAGATATGACATCTCTGTAACCTGCAACCGGATTTCTGACATGGCAGAGCT 707  
|||||  
Db 361 CTGGACAGATATGACATCTCTGTAACCTGCAACCGGATTTCTGACATGGCAGAGCT 420  
|||||  
QY 708 GTAATGTTAATGATATCCATTTTGGATTTTCTTTTCTTTTCTTTCTTTCTTTCTTTT 767  
|||||  
Db 421 GTAATGTTAATGATATCCATTTTGGATTTTCTTTTCTTTTCTTTCTTTCTTTCTTTT 480  
|||||  
QY 768 GAGGTAAATTTTTCAGTCTTCAAAAGTGGAAATACCTGG 806  
|||||  
Db 481 GAAGTAAATTTTTCAGTCTTCAAAAGTGGAAATACATGG 519  
|||||

RESULT 10  
BE824614/c  
LOCUS BE824614  
DEFINITION BE824614-3 NIA Mouse E13.5 VMB Dopamine cell cDNA Library Mus  
musculus cDNA clone C0502B03 3', mRNA sequence.  
ACCESSION BE824614  
VERSION BE824614.1 GI:10256848  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 587)  
REFERENCE Barrett, T., Xie, T., Piao, Y., Dillon-Carter, O., Kargul, G.J., Lim  
M.K., Chrest, F.J., Westo, R., Rowley, D.L., Juhaszova, M., Zhou, L.,  
Vavter, M.P., Becker, K.G., Cheadle, C., Wood, W.H. III, McCann, U.D.,  
Freed, W.J., Ko, M.S.H., Ricaurte, G., and Donovan, D.M.  
Genes expressed by midbrain dopamine neurons: a dopamine  
neuron-specific cDNA library and microarray used to examine changes  
in gene expression during development and methamphetamine  
neurotoxicity  
Unpublished (2000)  
Contact: George J. Kargul  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@igsun.grc.nia.nih.gov  
niaEST (<http://lgsun.grc.nia.nih.gov/cDNA/cDNA.html>)





digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Yulan Piao and Minoru.S. H. Ko."

```
BASE COUNT      194 a 101 c 86 g 210 t
ORIGIN
Query Match      23.1%; Score 435; DB 12; Length 591;
Best Local Similarity 86.8%; Pred. No. 2e-50;
Matches 515; Conservative 0; Mismatches 70; Indels 8; Gaps 3;

QY 1244 TTTATGTTTGGCCCAAGTGACCTTTTGTAGTAAATTAAGATTGTGTTTTTTAGTCATGCG 1303
DB 591 TCTATGTTTGGCCCAAGTGACCTTTTGTAGTAAATTAAGATTGTGTTTTTTAGTCATGCG 532
QY 1304 TTATGGACACACTATATTTCACCTCTTATATATGATCATCTACAGACAAAATTTCAAAA 1363
DB 531 TTATGGACACAGATATTCCACCTCTCTGTGATCATCTACAGACAAAATTTCAAAA 472
QY 1364 GGTCTTGTAAGATTAATGAAAGCGAGTTGTTTCTATAGTAGAGCTGATCCCGTCGCC 1423
DB 471 GGTCTTAAGAGTAGATGAAAGCGAGTTGTTTCCATAGTTGAAGCTGATCCCGTCGCC 412
QY 1424 TAATAATGCTGTATACAACTCTTGGATAGATCCCAAGAAACAAACAAAATTTACCTT 1483
DB 411 TAATAACGCTGTATACAACTCATGCTAGATCCCAAGAAACAAACAAAATTTACCTA 352
QY 1484 TGAAGATAGTGAATAAGAGAAAACGTTTAGTGCTCAGGTGTCACAGACTAGAGAAA 1543
DB 351 TGAAGACAGTGAATAAGAGAGAAATGTTTAGTACCTCAGGTGTCACAGACTAGAGAGA 292
QY 1544 AGTCFCAGTTTCCACCAATCCACATTCACAAATGAGTTTAAATTTAAATTTGAAACACTGA 1603
DB 291 AGTCFCAGCGCCACCAAT--ACATTCAGAGAGTTTAAATTTAAATTTGAAATGA 234
QY 1604 TATTACTGCCAATATAGAAAATPATTATTAAGTATGTTGTTGTAATTTTCAATG 1663
DB 233 ATTACTGCCAATGTTAGAGAAAACATCTAGGTATGTTGTTGTAATTTTCAATG 174
QY 1664 TGAATGCTATATAGATAGGTCATATATATATATATATATATATATATATATATATATAT 1719
DB 173 TGAATGCTATATAGATAGGTCATATATATATATATATATATATATATATATATATATAT 114
QY 1720 TGTTCATGCGCAGTTTGT--AAAGTACTATCATGTGTATATTTTGTCAATATATATGTC 1777
DB 113 TGTTCATGCGCAGTTTGT--AAAGTACTATCATGTGTATATTTTGTCAATATATATGTC 54
QY 1778 AACAGAAATATTCATGTAAGTCATATATTTTAAAGGAATAAATACATAGCCTT 1830
DB 53 ATCAGAAGATATCCATGTAAGTCATATATTTTCTAAAGAAATAAATATGAGCCTT 1
```

```
RESULT 12
BE824766/c
LOCUS BE824766 587 bp mRNA linear EST 21-SEP-2000
DEFINITION C0504E03-3 NIA Mouse E13.5 VMB Dopamine cell cDNA Library Mus
ACCESION BE824766
VERSION BE824766.1 GI:10257000
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
1 (bases 1 to 587)
REFERENCE Barrett,T., Xie,T., Piao,X., Dillon-Carter,O., Kargul,G.J., Lim
AUTHORS M.K., Chrest,F.J., Wersto,R., Rowley,D.L., Juhaszova,M., Zhou,L.,
Vavter,M.P., Becker,K.G., Chheadle,C., Wood,W.H. III, McCann,U.D.,
Freed,W.J., Ko,M.S.H., Ricaurte,G. and Donovan,D.M.
Genes expressed by midbrain dopamine neurons: a dopamine
neuron-specific cDNA library and microarray used to examine changes
```

JOURNAL  
COMMENT

in gene expression during development and methamphetamine neurotoxicity  
Unpublished (2000)  
Contact: George J. Kargul  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
Email: cdna@lgsun.grc.nia.nih.gov  
niaEST (http://lgsun.grc.nia.nih.gov/cDNA/cDNA.html)  
Plate: C0504 row: E column: 03  
Seq primer: -21M13 Forward  
High quality sequence stop: 587  
POLYA-Yes.

FEATURES  
Source

Location/Qualifiers  
1. 587  
/organism="Mus musculus"  
/strain="TH-beta-gal transgenic mouse"  
/db\_xref="niaEST:C0504E03-3"  
/db\_xref="taxon:10090"  
/clone="C0504E03"  
/clone.lib="NIA Mouse E13.5 VMB Dopamine cell cDNA Library"  
/sex="Male/Female"  
/dev\_stage="13.5dpc"  
/lab\_host="DH10B"  
/note="vector: pSPORT1 (Gibco/BRL Life Technology); Site 1: SalI; Site 2: NotI; Total RNAs were extracted from 3000 Dopamine cells (cell collected by Dr.Yanya Barrett). The double-stranded cDNA was synthesized by Gibco's kit with an oligo(dT) primer [NotI primer-adaptor from GibcoBRL] [5'-pgAGTAGTTCTAGATCGGCGGCCCTTTTTTTTTTTT-3'] from 0.9ug of total RNA. The double-stranded cDNAs were treated with T4 DNA polymerase and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker Li-Sal4 (include SalI sequence). The cDNAs were purified by phenol/chloroform and separated from free linkers by Centricon 100. Then, cDNAs were amplified by long-range high fidelity PCR using Takara's Ex Taq polymerase. Then, the cDNAs were purified by phenol/chloroform and by Centricon 100. The cDNAs were digested with SalI and NotI enzymes. Then, the cDNAs were size selected by Gibco's Size Fractionation Column. The cDNAs were cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by chemical method. The library was constructed by Yulan Piao and Minoru.S. H. Ko."

BASE COUNT 190 a 101 c 87 g 208 t 1 others  
ORIGIN

```
Query Match      22.9%; Score 431.6; DB 12; Length 587;
Best Local Similarity 86.8%; Pred. No. 5.7e-50;
Matches 511; Conservative 0; Mismatches 70; Indels 8; Gaps 3;

QY 1248 TGTTTAGCCCAAGTGACCTTTTGTAGTAAATTAAGATTGTGTTTTTTAGTCATGCTTAT 1307
DB 587 TGTTTAGCCCAAGTGACCTTTTGTAGTAAATTAAGATTGTGTTTTTTAGTCATGCTTAT 528
QY 1308 GGAACAACATATATTTTACCCTCTATATATATGATTCATCAGTACACAAAATTTCAAAGGTC 1367
DB 527 GGAACAACGATATTCACCTCTCTTGTATGATTCACACAAAATTTCAAAGGTC 468
QY 1368 TTGAAAAGTAAATGAAAAGCGAGTTGTTTCTATAGTAGAAGCTGATCCCTCGCTTAT 1427
DB 467 TTAAGAGTAGATGAAAAGCGAGTTGTTTCTATAGTAGAAGCTGATCCCTCGCTTAT 408
QY 1428 AATGCTGTAATACAACTCTTTGGATAGATCCCAAGAAACAAAATTTACCTTTGAA 1487
DB 407 AACGCTGTAATACAACTCTATGATAGATCTCTAAAGAAACAAAATTTACCTTATGAA 348
QY 1488 GATAGTGAATAAGAGAAAACGTTTGTAGTCTCAGGTGTCACAGACTAGAGAAAGTC 1547
DB 347 GACAGTGAATAAGAGAAAATGTTTGTAGTACTCAGGTGTCACAGACTAGAGAAAGTC 288
```







Db 102 TATCATGTGTATATTTTGTCAATATTATGTCCATCAGAGATATCCCATGTAAGTCATATT 43

QY 1806 TTTTAAGGAATAAATACATAGCCTTAAACA 1836

Db 42 TTCTAAAGAATAAATATGTAGCCTTAAAAA 12

Search completed: December 10, 2002, 09:44:13  
Job time : 2418 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: December 10, 2002, 07:34:42 ; Search time 4717 Seconds  
(without alignments)  
11605.338 Million cell updates/sec

Title: US-09-845-721-1  
Perfect score: 1881  
Sequence: 1 gttattttcctcaaaagaaa.....aaaaaaaaaaaaaaaaaaaaa 1881

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.in.\*
- 18: em.hum.\*
- 19: em.mu.\*
- 20: em.mu.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.in.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.man.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1881	100.0	1881	6	AX298123 Sequence
2	1881	100.0	1881	9	HSU66581 Human putat
3	1826.6	97.1	172533	9	AC002381 Human BAC
4	1625	86.4	2238	6	AR091359 Sequence
5	1388.8	73.8	83627	2	AC096185 Rattus no
6	535.8	28.5	76186	2	AC125020 Mus muscu
c 7	78.8	4.2	97683	2	AC116548 Dictyoste
c 8	76	4.0	201299	9	AC007158 Homo sapi
c 9	75.6	4.0	136098	9	AC006970 Homo sapi
c 10	75	4.0	93491	2	AC116967 Dictyoste
c 11	74.4	4.0	34980	6	AX344555 Sequence
c 12	73.2	3.9	106958	9	AL807813 Human DNA
c 13	72	3.9	201299	9	AC007158 Homo sapi
c 14	72	3.8	91723	9	AC010903 Homo sapi
c 15	71.8	3.8	25117	2	AC116102 Dictyoste
c 16	71.6	3.8	198431	9	AC109994 Homo sapi
c 17	71	3.8	11829	3	AE001425 Plasmodiu
c 18	71	3.8	191366	2	AC115351 Rattus no
c 19	70.6	3.8	2982	3	DDISGSPA M33862 Dictyosteli
c 20	70.4	3.7	141017	2	AC116962 Dictyoste
c 21	70.2	3.7	84805	2	AC116918 Dictyoste
c 22	70.2	3.7	155317	2	AC067973 Homo sapi
c 23	69	3.7	6007	3	AF474335 Dictyoste
c 24	69	3.7	118001	9	AC073081 Homo sapi
c 25	68.6	3.6	123280	2	AC117076 Dictyoste
c 26	68.6	3.6	206059	2	AC127383 Homo sapi
c 27	68.6	3.6	318221	2	PFMAL13P3 AL049184 Plasmodiu
c 28	68.4	3.6	149752	9	AC004616 Homo sapi
c 29	68.2	3.6	16343	2	AC006280 Plasmodiu
c 30	68.2	3.6	349980	6	AX344573 Sequence
c 31	68	3.6	133148	9	AC009435 Homo sapi
c 32	67.6	3.6	5487	6	AX346500 Sequence
c 33	67.6	3.6	176898	9	AC117569 Homo sapi
c 34	67.4	3.6	96183	9	AL161896 Human DNA
c 35	67.4	3.6	130349	9	AC011593 Homo sapi
c 36	67.2	3.6	131682	9	AL672277 Human DNA
c 37	67.2	3.6	198431	9	AC109994 Homo sapi
c 38	67.2	3.6	321003	2	PFMAL4P3 AL035476 Plasmodiu
c 39	67	3.6	2452	3	AF337815 Dictyoste
c 40	66.8	3.6	6047	2	AC115582 Dictyoste
c 41	66.8	3.6	40033	9	HS1110P6 AL049175 Human DNA
c 42	66.8	3.6	159942	9	AC025018 Homo sapi
c 43	66.8	3.6	187190	2	AC026898 Homo sapi
c 44	66.6	3.5	12029	3	AE001379 Plasmodiu
c 45	66.6	3.5	72243	9	AL731858 Human DNA

ALIGNMENTS

RESULT 1	AX298123	Sequence 1 from Patent WO0183550.	1881 bp	DNA	linear	PAT 26-NOV-2001
LOCUS	AX298123					
DEFINITION	AX298123					
ACCESSION	AX298123.1	GI:17128199				
VERSION						
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1					
AUTHORS	Brennan,J.C. and Hart,K.A.					
TITLE	Gpr22 modulators as appetite-control agents					
JOURNAL	Patent: WO 0183550-A 1 08-NOV-2001;					

FEATURES	source	Location/Qualifiers	1. .1881	288 g	627 t	0;
Astrazeneca AB (SE)						
BASE COUNT	649 a	317 c	288 g	627 t		
ORIGIN						
Query Match	100.0%	Score 1881;	DB 6;	Length 1881;		
Best Local Similarity	100.0%;	Pred. No. 1.3e-304;				
Matches 1881;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1	GTATTTCTTCAAGAGAAACACAAATTTCTTTATATCAAAAACAAATGCAAACTTGATG	60			
Db	1	GTATTTCTTCAAGAGAAACACAAATTTCTTTATATCAAAAACAAATGCAAACTTGATG	60			
QY	61	GTTCCTTAATCTACATTTTCTTATTAATAGTTTACAACTTAAAAATTAACCTAAGTACAC	120			
Db	61	GTTCCTTAATCTACATTTTCTTATTAATAGTTTACAACTTAAAAATTAACCTAAGTACAC	120			
QY	121	AATTGAAAGATTTTCTTTTACAAAGAACACGTTATACGTCATTTAAATTTGCCAAATA	180			
Db	121	AATTGAAAGATTTTCTTTTACAAAGAACACGTTATACGTCATTTAAATTTGCCAAATA	180			
QY	181	TCAATAGTTTATCTATTTTTCACCTTTCTAGGAAAAAACCACTGCTCCAAAAAGATGT	240			
Db	181	TCAATAGTTTATCTATTTTTCACCTTTCTAGGAAAAAACCACTGCTCCAAAAAGATGT	240			
QY	241	GTTTTTCCTCCATCTCGAAATCAACATGTCAGTCTGTAATCTAACATTTACAGTGGAGATG	300			
Db	241	GTTTTTCCTCCATCTCGAAATCAACATGTCAGTCTGTAATCTAACATTTACAGTGGAGATG	300			
QY	301	ACATTGATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTTCAAG	360			
Db	301	ACATTGATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTTCAAG	360			
QY	361	TGCTCTCACCGGATTTCTTATGTTAGAAATGTGTTGGGACTTTGGCAGCAACCTCACTG	420			
Db	361	TGCTCTCACCGGATTTCTTATGTTAGAAATGTGTTGGGACTTTGGCAGCAACCTCACTG	420			
QY	421	TATTTGGTACTTTACTGCATGAAATCCCACTTTAATCAACTCTGTCACTAACAATTATTACAA	480			
Db	421	TATTTGGTACTTTACTGCATGAAATCCCACTTTAATCAACTCTGTCACTAACAATTATTACAA	480			
QY	481	TGAATCTTCATGTACTTTGATTAATAATTTGTGGGATGTATTCCTCTCACTATACGTA	540			
Db	481	TGAATCTTCATGTACTTTGATTAATAATTTGTGGGATGTATTCCTCTCACTATACGTA	540			
QY	541	TCCCTCTGCTTTCACTGGAGAGTAACACTGCTCTCAATTTGCTGTTCCATGAGGCTTTGTG	600			
Db	541	TCCCTCTGCTTTCACTGGAGAGTAACACTGCTCTCAATTTGCTGTTCCATGAGGCTTTGTG	600			
QY	601	TATCTTTTGCAGTGTCTCAACAGCAATCAACGTTTTTGTATCATCATTGCGACAGATATG	660			
Db	601	TATCTTTTGCAGTGTCTCAACAGCAATCAACGTTTTTGTATCATCATTGCGACAGATATG	660			
QY	661	ACATCTCTGTAACACCTGCAACCGAATCTTGACAATGGCAGAGCTGTAATGTAATGA	720			
Db	661	ACATCTCTGTAACACCTGCAACCGAATCTTGACAATGGCAGAGCTGTAATGTAATGA	720			
QY	721	TATCCATTTGGAATTTTCTTTTCTCTTTCTCTGATTTCTTTTATTTAGGTAATTTTTT	780			
Db	721	TATCCATTTGGAATTTTCTTTTCTCTTTCTCTGATTTCTTTTATTTAGGTAATTTTTT	780			
QY	781	TCAGTCTTCAAGTGGAAATACCTGGGAAACAAAGACATTTTATGTGTCAGTACAAATG	840			
Db	781	TCAGTCTTCAAGTGGAAATACCTGGGAAACAAAGACATTTTATGTGTCAGTACAAATG	840			
QY	841	AATACTACACTGAACCTGGGAATGTATTCACCTGTTAGTACAGATCCCAATATCTTTT	900			
Db	841	AATACTACACTGAACCTGGGAATGTATTCACCTGTTAGTACAGATCCCAATATCTTTT	900			
QY	901	TCACTGTTGTAGTAAATGTTAATCACATACACCAAAATACCTCAGGCTCTTAAATATTCGAA	960			



```
QY 1441 ACAACTCTTGGATAGATCCCAAAAGAAACAAAAAATACCTTTGAAGATAGTGAATATA 1500
Db 1441 ACAACTCTTGGATAGATCCCAAAAGAAACAAAAAATACCTTTGAAGATAGTGAATATA 1500
QY 1501 GAGAAAACGTTTGTGCGCCAGGTTGTCACAGACTAGAGAAAAGTCTCAGTTTCACCAA 1560
Db 1501 GAGAAAACGTTTGTGCGCCAGGTTGTCACAGACTAGAGAAAAGTCTCAGTTTCACCAA 1560
QY 1561 ATCCACATTCAAATGAGTTTAAATTTAAATTTGTAATAAATGTAATAAATGTAATAAATGTAATAA 1620
Db 1561 ATCCACATTCAAATGAGTTTAAATTTAAATTTGTAATAAATGTAATAAATGTAATAAATGTAATAA 1620
QY 1621 AGAAAATATTTAAGTATGTTGTTAATTTGTAATAAATGTAATAAATGTAATAAATGTAATAA 1680
Db 1621 AGAAAATATTTAAGTATGTTGTTAATTTGTAATAAATGTAATAAATGTAATAAATGTAATAA 1680
QY 1681 AGGTCATATATATTCATATTCCTTCAATGATTAATGTTGTTGTTGTTGTTGTTGTTGTTGTTA 1740
Db 1681 AGGTCATATATATTCATATTCCTTCAATGATTAATGTTGTTGTTGTTGTTGTTGTTGTTA 1740
QY 1741 AGTACTATCATGTTGTTAATTTGTTCAATATATATGTTCCACAGAAAATATTCATGTTAAGTC 1800
Db 1741 AGTACTATCATGTTGTTAATTTGTTCAATATATATGTTCCACAGAAAATATTCATGTTAAGTC 1800
QY 1801 ATATTTTAAAGCAATTAATACATACAGCTTAACAGAGTGTATTAACCTTTAAATGTTAAATA 1860
Db 1801 ATATTTTAAAGCAATTAATACATACAGCTTAACAGAGTGTATTAACCTTTAAATGTTAAATA 1860
QY 1861 AAAAAAATATATATATTTGTTCAATATATATGTTCCACAGAAAATATTCATGTTAAGTC 1881
Db 1861 AAAAAAATATATATATTTGTTCAATATATATGTTCCACAGAAAATATTCATGTTAAGTC 1881
```

```
RESULT 3
AC002381 172533 bp DNA linear PRI 04-FEB-2000
LOCUS Human BAC clone CTB-20D2 from 7q22, complete sequence.
DEFINITION AC002381
ACCESSION AC002381
VERSION AC002381.1 GI:2275186
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
```

EuKaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```
REFERENCE
AUTHORS Gattung, S.
TITLE The sequence of H. sapiens BAC clone CTB-20D2
JOURNAL Unpublished
REFERENCE
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1997)
REFERENCE
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (04-FEB-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_RG020D02
-----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7> or send an E-mail to [egreen@nhgri.nih.gov](mailto:egreen@nhgri.nih.gov)

**SOURCE INFORMATION:**

This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).  
VECTOR: pBelOBAC11  
Selection: chloramphenicol

**NEIGHBORING SEQUENCE INFORMATION:**

The actual start of this clone is at base position 1 of CTB-20D2; actual end is at 172533 of CTB-20D2. The orientation of this clone is unknown.

This clone contains STS SWSS1826 (NID:gl113234) and SWSS869 (NID:g23340).

FEATURES	Source	Location/Qualifiers
	1. .172533	/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="7"
		/map="7q22"
repeat_region		/clone_lib="CTB-20D2"
		/clone="CTB-978SK-B"
		complement(1060..1352)
misc_feature		/rpt_family="ALU"
		complement(1353..1438)
		/note="match to human EST AA225239 (NID:g1846547)"
misc_feature		nc22d09.r1
		complement(1353..1438)
		/note="match to human EST AA225630 (NID:g1846956)"
misc_feature		nc08a05.r1
		1353..1431
		/note="match to human EST AA225171 (NID:g1846460)"
misc_feature		nc22d09.s1
		1353..1420
		/note="match to human EST AA225498 (NID:g1846862)"
repeat_region		nc08a05.s1
		3499..3596
repeat_region		/rpt_family="LI"
		complement(3803..4141)
repeat_region		/rpt_family="MER"
		4164..4203
repeat_region		/rpt_family="LI"
		4418..4709
repeat_region		/rpt_family="ALU"
		4771..4797
repeat_region		/rpt_family="LI"
		complement(4806..5036)
repeat_region		/rpt_family="ALU"
		complement(5050..5182)
repeat_region		/rpt_family="ALU"
		5249..5281



```
repeat_region /rpt_family="L1"
5317..5493
/rpt_family="L1"
7526..7555
/rpt_family="L1"
complement(8884..9083)
/rpt_family="L1"
9114..9199
/notes="match to human EST R05512 (NID:g756132) ye92e05.s1"
misc_feature complement(9688..9980)
misc_feature /notes="match to human EST R05619 (NID:g756239) ye92e05.r1"
repeat_region 12598..12886
repeat_region /rpt_family="ALU"
14708..14768
/rpt_family="L1"
complement(15481..15539)
/rpt_family="L1"
complement(15834..15947)
/rpt_family="L1"
/notes="match to human EST AA094108 (NID:g1639701)"
repeat_region complement(15948..16239)
/rpt_family="ALU"
misc_feature complement(18000..18131)
/notes="match to human EST AAL188905 (NID:g1775949)
2979f02.r1"
misc_feature complement(18000..18131)
/notes="match to human EST AA206526 (NID:g1801907)
2d58c01.r1"
repeat_region complement(18508..18770)
/rpt_family="ALU"
repeat_region 19013..19068
/rpt_family="L1"
repeat_region complement(20234..20523)
/rpt_family="ALU"
repeat_region 21100..21124
/rpt_family="L1"
repeat_region 21172..21542
/rpt_family="L1"
complement(21174..21536)
/rpt_family="L1"
21731..21764
/rpt_family="L1"
23854..23898
/rpt_family="L1"
complement(24280..24523)
/rpt_family="ALU"
repeat_region 24969..24996
/rpt_family="L1"
complement(25073..25365)
/rpt_family="ALU"
repeat_region complement(25967..26258)
/rpt_family="ALU"
26790..27002
/rpt_family="L1"
27007..27161
/rpt_family="ALU"
27171..27273
/rpt_family="ALU"
27593..27650
/rpt_family="L1"
27831..28035
/rpt_family="ALU"
28177..28205
/rpt_family="L1"
28909..28986
/rpt_family="L1"
29056..29327
/rpt_family="L1"
29372..29429
/rpt_family="L1"
29489..29780
/rpt_family="ALU"
31289..31368
/rpt_family="THE"
```

```
repeat_region 31384..31694
/rpt_family="THE"
repeat_region complement(34785..34956)
/rpt_family="L1"
repeat_region complement(34958..35069)
/rpt_family="ALU"
repeat_region complement(35451..36208)
/rpt_family="L1"
repeat_region 36175..36203
/rpt_family="L1"
repeat_region 36479..36545
/rpt_family="L1"
repeat_region complement(37143..37236)
/rpt_family="ALU"
repeat_region complement(37259..37450)
/rpt_family="ALU"

Query Match 97.1%; Score 1826.6; DB 9; Length 172533;
Best Local Similarity 99.4%; Pred. No. 5.5e-296;
Matches 1854; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

QY 1 GTTATTTCTTCAAAAGGAAACACAAATTTCTTTTATATCAAAACAATGCAAACTTGATG 60
Db 79324 GTTATTTCTTCAAAAGGAAACACAAATTTCTTTTATATCAAAACAATGCAAACTTGATG 79383
QY 61 GTTCTTAATTTCTACATTTTCTATTAATAGTTTACAAACTTAAATAAAGTACAC 120
Db 79384 GTTCTTAATTTCTACATTTTCTATTAATAGTTTACAAACTTAAATAAAGTACAC 79443
QY 121 AATTGAAAGATTTTCTTCTTCAAAAGAACACGCTTATAGTCATTAAATTTGCCAATA 180
Db 79444 AATTGAAAGA-TTTTCTTCTTCAAAAGAACACGCTTATAGTCATTAAATTTGCCAATA 79502
QY 181 TCAATAGTTTATTTTCACTTTCTAGGAAACCAACTGCTCCAAAAAGATGT 240
Db 79503 TCAATAGTTTATTTTCTATTTCTAGGAAACCAACTGCTCCAAAAAGATGT 79562
QY 241 GTTTTCTCCATCTCGGAATCAACATGAGTCTGAATCTAATCTACAGTGGAGATG 300
Db 79563 GTTTTCTCCATCTCGGAATCAACATGAGTCTGAATCTAATCTACAGTGGAGATG 79622
QY 301 ACATTGATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAGCTTCAAG 360
Db 79623 ACATTGATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAGCTTCAAG 79682
QY 361 TGTCTCTCACCGGATTTCTTATGTTAGAAAATGTGTTGGACTTGGCAGCAACCTCACTG 420
Db 79683 TGTCTCTCACCGGATTTCTTATGTTAGAAAATGTGTTGGACTTGGCAGCAACCTCACTG 79742
QY 421 TATTTGTTACTTACTGTCATGAAATCCAACTTAATCAACTCTGTCAGTAAACATTTATCAA 480
Db 79743 TATTTGTTACTTACTGTCATGAAATCCAACTTAATCAACTCTGTCAGTAAACATTTATCAA 79802
QY 481 TGAATCTTCATGTTACTTGTATTAATTTGTGTGGAGTATTCTCTTAAGTATTAGTTA 540
Db 79803 TGAATCTTCATGTTACTTGTATTAATTTGTGTGGAGTATTCTCTTAAGTATTAGTTA 79862
QY 541 TCCTTCTGCTTTCACTGGAGAGTAAACACTGCTCTCATTTGCTTTTCCATGAGGCTTTGTG 600
Db 79863 TCCTTCTGCTTTCACTGGAGAGTAAACACTGCTCTCATTTGCTTTTCCATGAGGCTTTGTG 79922
QY 601 TATCTTTTGAAGTGTCTCAACGCAATCAACGTTTTTGTCTATCCTTTTGACAGATATG 660
Db 79923 TATCTTTTGAAGTGTCTCAACGCAATCAACGTTTTTGTCTATCCTTTTGACAGATATG 79982
QY 661 ACATCTCTGTAACACCTGCAACCGAATCTTGACAATGGGCGAGAGCTGTAATGTAATGA 720
Db 79983 ACATCTCTGTAACACCTGCAACCGAATCTTGACAATGGGCGAGAGCTGTAATGTAATGA 80042
QY 721 TATCCATTTGGATTTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 780
Db 80043 TATCCATTTGGATTTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 80102
```

QY 781 TCAGTCTTCAAGTGGAAATACCTGGGAAACAACAGACACTTTTATGTGTGCTAGTACAAATG 840  
Db 80103 TCAGTCTTCAAGTGGAAATACCTGGGAAACAACAGACACTTTTATGTGTGCTAGTACAAATG 80162  
QY 841 AATACTACACTGAATGGGAATGTATATACCTGTGTAGTACAGATCCCAATATTTCTTTT 900  
Db 80163 AATACTACACTGAATGGGAATGTATATACCTGTGTAGTACAGATCCCAATATTTCTTTT 80222  
QY 901 TCACCTGTGTAGTAAATGTAATACATACACAAAATACTTCAGGCTCTTAATATTCGAA 960  
Db 80223 TCACCTGTGTAGTAAATGTAATACATACACAAAATACTTCAGGCTCTTAATATTCGAA 80282  
QY 961 TAGGCACAGATTTTCAACAGGGCAGAGAGAAAGCAGAAAGAAAGACAAATTTCTC 1020  
Db 80283 TAGGCACAGATTTTCAACAGGGCAGAGAGAAAGCAGAAAGAAAGACAAATTTCTC 80342  
QY 1021 TAACACACAAACATCAGGCTACAGACATGTCAACAAGCAGTGGTGGGAGAAATGTAGTCT 1080  
Db 80343 TAACACACAAACATCAGGCTACAGACATGTCAACAAGCAGTGGTGGGAGAAATGTAGTCT 80402  
QY 1081 TTGGTGTAAAGAACTTCAGTTCCTGTAAATATTCGCCCTCCGGCAGCTGTGAAACGACACC 1140  
Db 80403 TTGGTGTAAAGAACTTCAGTTCCTGTAAATATTCGCCCTCCGGCAGCTGTGAAACGACACC 80462  
QY 1141 GTGAACGAGGAGAAAGACAAAGAGAGTCTTCAGGATGTCTTTATGTATTTCTACAT 1200  
Db 80463 GTGAACGAGGAGAAAGACAAAGAGAGTCTTCAGGATGTCTTTATGTATTTCTACAT 80522  
QY 1201 TTCTTCTCTGCTGGACACCAATTTCTGTTTTAAATACCAACCAATTTATGTTTTAGGCCCAA 1260  
Db 80523 TTCTTCTCTGCTGGACACCAATTTCTGTTTTAAATACCAACCAATTTATGTTTTAGGCCCAA 80582  
QY 1261 GTGACCTTTTGTAGTAAATTAAGATTTGTGTTTTTATGTCATGSCCTTATGGAACAATAT 1320  
Db 80583 GTGACCTTTTGTAGTAAATTAAGATTTGTGTTTTTATGTCATGSCCTTATGGAACAATAT 80642  
QY 1321 TTCACCTCTATATATGATTCATCAGTACACAAATTTTCAAAAGCTCTGAAAGTAAA 1380  
Db 80643 TTCACCTCTATATATGATTCATCAGTACACAAATTTTCAAAAGCTCTGAAAGTAAA 80702  
QY 1381 TGAAGAAAGCGAGTTGTTCTATAGTAGAGCTGATCCCTGCTTAATATGCTGTAATAC 1440  
Db 80703 TGAAGAAAGCGAGTTGTTCTATAGTAGAGCTGATCCCTGCTTAATATGCTGTAATAC 80762  
QY 1441 ACACTCTTGGATAGATCCCAAGAAACAAACAAATTAACCTTTTGAAGATAGTGAATAA 1500  
Db 80763 ACACTCTTGGATAGATCCCAAGAAACAAACAAATTAACCTTTTGAAGATAGTGAATAA 80822  
QY 1501 GAGAAAGAGTTTGTGCTCAGTGTGTACAGACTAGAGAAAGTCTCAGTTTCACCAA 1560  
Db 80823 GAGAAAGAGTTTGTGCTCAGTGTGTACAGACTAGAGAAAGTCTCAGTTTCACCAA 80882  
QY 1561 ATCCACATTCAAATCAGTTTAAATTTAAATTTGTAAGTAAATTTTCAATGCTGAATATAGAT 1620  
Db 80883 ATCCACATTCAAATCAGTTTAAATTTAAATTTGTAAGTAAATTTTCAATGCTGAATATAGAT 80942  
QY 1621 AGAAAAATATTTTAAAGTATGTTGTTAAATTTTCAATGCTGAATATAGAT 1680  
Db 80943 AGAAAAATATTTTAAAGTATGTTGTTAAATTTTCAATGCTGAATATAGAT 81001  
QY 1681 AGGTATATATATCAATTTCTCAATTTCTCAATTTATGTTGATGATGATGATGATGATGATGAT 1740  
Db 81002 AGGTATATATATCAATTTCTCAATTTCTCAATTTATGTTGATGATGATGATGATGATGATGAT 81061  
QY 1741 AGTACTATCATGTGTATATTTTGTCAATATTTGTCACACAGAAATATTCATGTAAGTC 1800  
Db 81062 AGTACTATCATGTGTATATTTTGTCAATATTTGTCACACAGAAATATTCATGTAAGTC 81121  
QY 1801 ATATTTTTTAAAGGAATAATACATAGCCTTAAACAGTGTATACTTTTAAATGTAAAAA 1860  
Db 81122 ATATTTTTTAAAGGAATAATACATAGCCTTAAACAGTGTATACTTTTAAATGTAAAAA 81181  
QY 1861 AAAAA 1865

Db 81182 ACATA 81186  
RESULT 4  
AR091359  
LOCUS AR091359 2238 bp DNA linear PAT 07-SEP-2000  
DEFINITION Sequence 2 from patent US 5994097.  
ACCESSION AR091359  
VERSION AR091359.1 GI:10018114  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2238)  
AUTHORS Lal, P., Guegler, K.J., Shah, P. and Corley, N.C.  
TITLE Polynucleotide encoding human G-protein coupled receptor  
JOURNAL Patent: US 5994097-A 2 30-NOV-1999;  
FEATURES  
Location/Qualifiers  
1..2238  
source  
BASE COUNT 773 a 385 c 348 g 732 t  
ORIGIN  
Query Match 86.4%; Score 1625; DB 6; Length 2238;  
Best Local Similarity 97.1%; Pred. No. 6.7e-262;  
Matches 1677; Conservative 0; Mismatches 45; Indels 5; Gaps 2;  
QY 139 TCTTACAAGAACACGTTTATACGTCATTTAAATTCGCAAAATATCAAAATAGTTTATTTCTAT 198  
Db 28 TATTGGAGAGAAAGCAAGAATTCCTCCAGACATATGAATAAACAACAGACTACTTC--- 84  
QY 199 TTTCACTTTCTAGGGAAAAAACAACCTGCTCCAAAAGAATGTGTTTTCTCCCATTTCTGG 258  
Db 85 -ACTGTGTGTCAGGAAAAAACAACCTGCTCCAAAAGAATGTGTTTTCTCCCATTTCTGG 143  
QY 259 AAATCAACATGAGCTGTAATCTAACATTTACAGTGGAGATGACATTTGATGATCAACA 318  
Db 144 AAATCAACATGAGCTGTAATCTAACATTTACAGTGGAGATGACATTTGATGATCAACA 203  
QY 319 CCAATATGTACCAACACATCATATCCGTTAAGCTTTCAAGTGTCTCTCACGGATTTTC 378  
Db 204 CCAATATGTACCAACACATCATATCCGTTAAGCTTTCAAGTGTCTCTCACGGATTTTC 263  
QY 379 TTATGTAGAAATTTGTTGGAGTTGGCAGCAACCTCACTGTTATGGTACTTTTACTGCA 438  
Db 264 TTATGTAGAAATTTTTTGGACTTTGGCAGCAACCTCACTGTTATGGTACTTTACTGCA 323  
QY 439 TGAATCCAACTTAATCAACTCTGTCAGTAACATTTATTAACAATGAAATCTTCATGACTTG 498  
Db 324 TGAATCCAACTTAATCAACTCTGTCAGTAACATTTATTAACAATGAAATCTTCATGACTTG 383  
QY 499 ATGTAATATTTGTGGGATGTTTCCCTCAACTATAGTATAGTATCTCTGCTTCTACTGG 558  
Db 384 ATGTAATATTTGTGGGATGTTTCCCTCAACTATAGTATAGTATCTCTGCTTCTACTGG 443  
QY 559 AGAGTAACACTGCTCTCAITTTGCTGTTTCCATGAGGCTTTGTATCTTTTGAAGTGTCT 618  
Db 444 AGAGTAACACTGCTCTCAITTTGCTGTTTCCATGAGGCTTTGTATCTTTTGAAGTGTCT 503  
QY 619 CAACAGCAATCAACCTTTTGTCTATCAGTTTGGACAGATATGACATCTCTGTAACCTTG 678  
Db 504 CAACAGCAATCAACCTTTTGTCTATCAGTTTGGACAGATATGACATCTCTGTAACCTTG 563  
QY 679 CAACCGGAATCTGCAATGGCGAGCTGTAATGTTAATGATATCCATTTGGATTTTTT 738  
Db 564 CAACCGGAATCTGCAATGGCGAGCTGTAATGTTAATGATATCCATTTGGATTTTTT 623  
QY 739 CTTTCTTCTCTTCTGATTTCTTTTATGAGTAAATTTTTCAGTCTTCTCAAGTGGAA 798  
Db 624 CTTTCTTCTCTTCTGATTTCTTTTATGAGTAAATTTTTCAGTCTTCTCAAGTGGAA 683  
QY 799 ATACCTGGAAAAACAAGACACTTTTATGTGTCAGTACAAATGAATACTACTGAACTGG 858

Db 684 ATACCTGGGAAACAGACACACTTTATGTGTCAAGTACAAATGAATACTACATGAACGTGG 743  
|||||  
QY 859 GAATGTATTATCACCTGTTAGTACAGATCCCAATATCTTTTTCACCTGTTGTAGTAAATGT 918  
|||||  
Db 744 GAATGTATTATCACCTGTTAGTACAGATCCCAATATCTTTTTCACCTGTTGTAGTAAATGT 803  
|||||  
QY 919 TAATCATATACACCAAAATACTTTCAGCTCTTAATATTCGAATAGGCACAAATTTTCAA 978  
|||||  
Db 804 TAATCATATACACCAAAATACTTTCAGCTCTTAATATTCGAATAGGCACAAATTTTCAA 863  
|||||  
QY 979 CAGGCGAAGAAAGAAAGCAAGAAAGAAAGCAATTTCTTAACACACAAATCATGAGG 1038  
|||||  
Db 864 CAGGCGAAGAAAGAAAGCAAGAAAGAAAGCAATTTCTTAACACACAAATCATGAGG 923  
|||||  
QY 1039 CTACAGACATGTCACAAAGCAGTGGTGGGAGAAATGTAGTCTTTTGGTGTGAAGAACTTCAG 1098  
|||||  
Db 924 CTACAGACATGTCACAAAGCAGTGGTGGGAGAAATGTAGTCTTTTGGTGTGAAGAACTTCAG 983  
|||||  
QY 1099 TTTCTGTAATTAATGCGCTCCGGGAGCTGTGAACGACACCGTGAACGAGGAGAAAGAC 1158  
|||||  
Db 984 TTTCTGTAATTAATGCGCTCCGGGAGCTGTGAACGACACCGTGAACGAGGAGAAAGAC 1043  
|||||  
QY 1159 AAAGAGAGCTTCAGGATCTTTATTGATTTATTTCTACATTTCTTCTCTGCTGGACAC 1218  
|||||  
Db 1044 TAAGAGAGCTTCAGGATGCTTTATTGATTTATTTCTACATTTCTTCTCTGCTGGACAC 1103  
|||||  
QY 1219 CAATTTCTGTTTAAATPACCACCATTTTATGTTTATGAGCCCAAGTGACCTTTTATAGTAAAT 1278  
|||||  
Db 1104 CAATTTCTGTTTAAATPACCACCATTTTATGTTTATGAGCCCAAGTGACCTTTTATAGTAAAT 1163  
|||||  
QY 1279 TAAGATTGCTTTTATGCTATGCTATGGAACAATATATTTCACCCCTCTATTATATG 1338  
|||||  
Db 1164 TAAGATTGCTTTTATGCTATGCTATGGAACAATATATTTCACCCCTCTATTATATG 1223  
|||||  
QY 1339 CATTCACACACAAAATTTTCAAGGCTTTGAAAGCTTTGAAAGCTTAAATGAAAGGAGCTTCTTT 1398  
|||||  
Db 1224 CATTCACACACAAAATTTTCAAGGCTTTGAAAGCTTTGAAAGCTTAAATGAAAGGAGCTTCTTT 1283  
|||||  
QY 1399 CTATAGTAGAGCTGATCCCTGCTTAATATGCTGTGAATACACAACTCTTGGATAGATC 1458  
|||||  
Db 1284 CTATAGTAGAGCTGATCCCTGCTTAATATGCTGTGAATACACAACTCTTGGATAGATC 1343  
|||||  
QY 1459 CCAAAAGAACACAAAATTTACCTTTCAAGATAGTGAATTAAGAGAAAACGTTTAGTGC 1518  
|||||  
Db 1344 CTAAGAGAAAACAAAATTTACCTTTGAAGATAGTGAATTAAGAGAAAATGTTTAGTGC 1403  
|||||  
QY 1519 CTCAGGTTGTACAGACTAGAGAAAAGCTTCAGTTTCCACCAATCCCAATCCCAATCAGT 1578  
|||||  
Db 1404 CTCAGGTTGTACAGACTAGAGAAAAGCTTCAGTTTCCACCAATCCCAATCAGT 1463  
|||||  
QY 1579 TTTAAATTTAAATTTGTAAGAACTGATATTACTGCAAAATATAAGAAAATATTTTAAAGTA 1638  
|||||  
Db 1464 TTTAAATTTAAATTTGTAAGAACTGATATTACTGCAAAATATAAGAAAATATTTTAAAGTA 1523  
|||||  
QY 1639 TTGTTATGTTGTAATTTTCAATGCTCAATGCTAATTTAGATAGGTCATATATTTCAAT 1698  
|||||  
Db 1524 TTGTTATGTTGTAATTTTCAATGCTG-ATGTCATTTAGATAGGTCATATATTTCAAT 1582  
|||||  
QY 1699 TTTCTCATTTACTTAATGCTATTTGTTGCTAGCGCTTTGTTAAAGTACTATCATGCTGATA 1758  
|||||  
Db 1583 TTTCTCATTTACTTAATGCTATTTGTTGCTAGCGCTTTGTTAAAGTACTATCATGCTGATA 1642  
|||||  
QY 1759 TTTTGTCAATATTATGTCACACGAAATATTTCATGTAAAGTCATATTTTAAAGGAATAA 1818  
|||||  
Db 1643 TTTTGTCAATATTATGTCACACGAAATATTTCATGTAAAGTCATATTTTAAAGGAATAA 1702  
|||||  
QY 1819 ATACATAGCCTTAAACAGGTATATACTTTAAATGTAAAGAAAAA 1865  
|||||  
Db 1703 ATACATAGCCTTAAACAGGTATATACTTTAAATGTAAAGTAACTGACATA 1749  
|||||

RESULT 5

AC096185  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AC096185 83627 bp DNA linear HTG 11-JUL-2002  
Rattus norvegicus clone CH230-11G1, \*\*\* SEQUENCING IN PROGRESS \*\*\*,  
43 unordered pieces.  
AC096185  
HTG: HTGS\_PHASE1.  
Norway rat.  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 83627)  
Kuzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbaja,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyie,M.D., Dathorne,S.R., David,R.,  
Davilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Donthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisseged,H.,  
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogih,M., Okwuonu,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,  
Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,  
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansey,J., Taylor,C., Taylor,T., Telrod,B., Thomas,N., Thomas,S.,  
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.  
Direct Submission  
Unpublished  
2 (bases 1 to 83627)  
Worley,K.C.  
Direct Submission  
Submitted (11-Sep-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 83627)  
Worley,K.C.  
Direct Submission  
Submitted (11-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 10, 2002 this sequence version replaced gi:17943871.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

----- Project Information  
Center project name: GBPL  
Center clone name: CH230-11G1  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 31803 bases at least Q40  
Consensus quality: 35840 bases at least Q30  
Consensus quality: 39549 bases at least Q20  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 43 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1247: contig of 1247 bp in length  
\* 1248: gap of unknown length  
\* 1348: contig of 1062 bp in length  
\* 2410: gap of unknown length  
\* 2510: contig of 1532 bp in length  
\* 4042: gap of unknown length  
\* 4142: contig of 1365 bp in length  
\* 5507: gap of unknown length  
\* 5606: gap of unknown length  
\* 7018: contig of 1412 bp in length  
\* 7019: gap of unknown length  
\* 7119: contig of 1413 bp in length  
\* 8532: gap of unknown length  
\* 8632: contig of 1117 bp in length  
\* 9749: gap of unknown length  
\* 9849: contig of 1131 bp in length  
\* 10980: gap of unknown length  
\* 11079: gap of unknown length  
\* 12217: contig of 1138 bp in length  
\* 12318: gap of unknown length  
\* 12318: contig of 1427 bp in length  
\* 13745: gap of unknown length  
\* 13845: contig of 1908 bp in length  
\* 15752: gap of unknown length  
\* 15853: contig of 1444 bp in length  
\* 17297: gap of unknown length  
\* 17397: contig of 1400 bp in length  
\* 18797: gap of unknown length  
\* 20199: contig of 1302 bp in length  
\* 20299: gap of unknown length  
\* 22039: contig of 1740 bp in length  
\* 22139: gap of unknown length  
\* 22139: contig of 1762 bp in length  
\* 23901: gap of unknown length  
\* 24001: contig of 1040 bp in length  
\* 25041: gap of unknown length  
\* 25141: contig of 1317 bp in length  
\* 26557: gap of unknown length  
\* 26558: contig of 1140 bp in length  
\* 27697: gap of unknown length  
\* 27798: contig of 1374 bp in length  
\* 29172: gap of unknown length  
\* 29272: contig of 1050 bp in length  
\* 30322: gap of unknown length  
\* 30422: contig of 1471 bp in length  
\* 31893: gap of unknown length  
\* 31993: contig of 2002 bp in length  
\* 33995: gap of unknown length  
\* 34095: contig of 2007 bp in length  
\* 36102: gap of unknown length  
\* 36202: contig of 1196 bp in length  
\* 37398: gap of unknown length  
\* 37498: contig of 1798 bp in length  
\* 39296: gap of unknown length

39396 41152: contig of 1757 bp in length  
41153: gap of unknown length  
41253: contig of 2813 bp in length  
44066: gap of unknown length  
44166: contig of 1711 bp in length  
45877: gap of unknown length  
45977: contig of 1876 bp in length  
47853: gap of unknown length  
47953: contig of 2020 bp in length  
49973: gap of unknown length  
50073: contig of 2729 bp in length  
52901: gap of unknown length  
52902: contig of 2474 bp in length  
55375: gap of unknown length  
55476: contig of 2569 bp in length  
58144: gap of unknown length  
58145: contig of 2032 bp in length  
60177: gap of unknown length  
60277: contig of 2516 bp in length  
62793: gap of unknown length  
62893: contig of 2706 bp in length  
65599: gap of unknown length  
65699: contig of 2142 bp in length  
67841: gap of unknown length  
67841: contig of 2646 bp in length  
70587: gap of unknown length  
70587: contig of 2819 bp in length  
73506: gap of unknown length  
73606: contig of 2531 bp in length  
76137: gap of unknown length  
76237: contig of 3935 bp in length  
80172: gap of unknown length  
80272: contig of 3356 bp in length.  
80272: Location/Qualifiers  
1.83627  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-11G1"  
BASE COUNT 23279 a 15508 c 16123 g 24487 t 4230 others  
ORIGIN  
Query Match 73.8%; Score 1388.8; DB 2; Length 83627;  
Best Local Similarity 86.8%; Pred. No. 7.9e-232;  
Matches 1622; Conservative 0; Mismatches 232; Indels 14; Gaps 8;  
QY 1 GTATTCTTCAAGAGAAACACAAATTTCTTTATATCAAAACAAATCAAACTTGATG 60  
DB 60431 GTATTCTTCAAAATACAA-ATTTCCCTTTACCCAAACAAATCGACCCCTCATG 60489  
QY 61 GTTCTAATTTACATTTTCTAATTAATAGTTTACAACTTAAAAATTAACCTAAGTACAC 120  
DB 60490 TTTCTCAATTTACATTTTGCATGACGAGTTTACAAAATACAAAATAACATACACAAT 60549  
QY 121 AATTCAAGATTTTCTTTTACAAAGAACACAGTTATACGTCATTTAAATTTGCCAAATA 180  
DB 60550 AA--AAAGATATCTTGTGTTTACAAAGAACATCGTGTCTTATATCATCTATAAAGTCGC 60607  
QY 181 TCAATATAGTTTATCTTCTTCTTCTAGGAGAAACCAACCTGCTCCAAAGATCT 240  
DB 60608 AAACATCAAAATATCTTATTTGATTTTCTAGGG-GAATACCAACCTGCTCCAAAGATCT 60666  
QY 241 GTTTTCTCCCAATCTCGAAATCAACATGACGTCGTAATCTAACATTTACAGTCCGAGATG 300  
DB 60667 GTTTTCTCCCTGTTCTGGAATCAACATGACGTCGTAATCAACATGACGTTGCGAGATG 60726  
QY 301 ACATTTGATGACATCAACACCAATATGTACCAACCACTATCATATCGGTTAAGCTTTCAAG 360  
DB 60727 ACATTTGAGGACATCGATACCAATATGTACCAACCACTGTCATACCAATTAAGCTTTCAAG 60786  
QY 361 TGTCTCTCACCGGATTTCTTATGTTAGAAATGTGTTGGAGCTTGGCAGCAACCTCACATG 420  
DB 60787 TGTCTCTCACGAGATTTCTCTATGTTAGAAATGTGTTGGGCTTGGTAGCAACCTTACCG 60846



Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submission@genome.wi.mit.edu](mailto:sequence_submission@genome.wi.mit.edu)

----- Project Information

Center project name: L25186

Center clone name: 181\_E\_5

\* NOTE: This record contains 91 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
\* 691 790: contig of 690 bp in length  
\* 791 790: gap of 100 bp  
\* 1551 1551: contig of 761 bp in length  
\* 1552 1651: gap of 100 bp  
\* 1652 2395: contig of 744 bp in length  
\* 2396 2495: gap of 100 bp  
\* 2496 3216: contig of 721 bp in length  
\* 3217 3316: gap of 100 bp  
\* 3317 4046: contig of 730 bp in length  
\* 4047 4146: gap of 100 bp  
\* 4147 4877: contig of 731 bp in length  
\* 4878 4977: gap of 100 bp  
\* 4978 5716: contig of 739 bp in length  
\* 5717 5816: gap of 100 bp  
\* 5817 6556: contig of 740 bp in length  
\* 6557 6656: gap of 100 bp  
\* 6657 7387: contig of 731 bp in length  
\* 7388 7487: gap of 100 bp  
\* 7488 8245: contig of 758 bp in length  
\* 8246 8345: gap of 100 bp  
\* 8346 9066: contig of 721 bp in length  
\* 9067 9166: gap of 100 bp  
\* 9167 9880: contig of 714 bp in length  
\* 9881 9980: gap of 100 bp  
\* 9981 10728: contig of 748 bp in length  
\* 10729 10828: gap of 100 bp  
\* 10829 11539: contig of 711 bp in length  
\* 11540 11639: gap of 100 bp  
\* 11640 12371: contig of 732 bp in length  
\* 12372 12471: gap of 100 bp  
\* 12472 13190: contig of 719 bp in length  
\* 13191 13290: gap of 100 bp  
\* 13291 14034: contig of 744 bp in length  
\* 14035 14134: gap of 100 bp  
\* 14135 14890: contig of 756 bp in length  
\* 14891 14990: gap of 100 bp  
\* 14991 15717: contig of 727 bp in length  
\* 15718 15817: gap of 100 bp  
\* 15818 16569: contig of 752 bp in length  
\* 16570 16669: gap of 100 bp  
\* 16670 17396: contig of 727 bp in length  
\* 17397 17496: gap of 100 bp  
\* 17497 18249: contig of 753 bp in length  
\* 18250 18349: gap of 100 bp  
\* 18350 19051: contig of 702 bp in length  
\* 19052 19151: gap of 100 bp  
\* 19152 19903: contig of 752 bp in length  
\* 19904 20003: gap of 100 bp  
\* 20004 20736: contig of 733 bp in length  
\* 20737 20836: gap of 100 bp  
\* 20837 21567: contig of 731 bp in length  
\* 21568 21667: gap of 100 bp  
\* 21668 22389: contig of 722 bp in length  
\* 22390 22489: gap of 100 bp

\* 22490 23215: contig of 726 bp in length  
\* 23216 23315: gap of 100 bp  
\* 23316 24044: contig of 729 bp in length  
\* 24045 24144: gap of 100 bp  
\* 24145 24885: contig of 741 bp in length  
\* 24886 24985: gap of 100 bp  
\* 24986 25745: contig of 760 bp in length  
\* 25746 25845: gap of 100 bp  
\* 25846 26702: contig of 757 bp in length  
\* 26703 27460: contig of 758 bp in length  
\* 27461 27560: gap of 100 bp  
\* 27561 28320: contig of 760 bp in length  
\* 28321 28420: gap of 100 bp  
\* 28421 29134: contig of 714 bp in length  
\* 29135 29234: gap of 100 bp  
\* 29235 29988: contig of 754 bp in length  
\* 29989 30088: gap of 100 bp  
\* 30089 30801: contig of 713 bp in length  
\* 30802 30901: gap of 100 bp  
\* 30902 31622: contig of 721 bp in length  
\* 31623 31722: gap of 100 bp  
\* 31723 32440: contig of 718 bp in length  
\* 32441 32540: gap of 100 bp  
\* 32541 33265: contig of 725 bp in length  
\* 33266 33365: gap of 100 bp  
\* 33366 34119: contig of 754 bp in length  
\* 34120 34219: gap of 100 bp  
\* 34220 34973: contig of 754 bp in length  
\* 34974 35073: gap of 100 bp  
\* 35074 35834: contig of 761 bp in length  
\* 35835 35934: gap of 100 bp  
\* 35935 36689: contig of 755 bp in length  
\* 36690 36789: gap of 100 bp  
\* 36790 37540: contig of 751 bp in length  
\* 37541 37640: gap of 100 bp  
\* 37641 38391: contig of 751 bp in length  
\* 38392 38491: gap of 100 bp  
\* 38492 39239: contig of 748 bp in length  
\* 39240 39339: gap of 100 bp  
\* 39340 40095: contig of 756 bp in length  
\* 40096 40195: gap of 100 bp  
\* 40196 40890: contig of 695 bp in length  
\* 40891 40990: gap of 100 bp  
\* 40991 41719: contig of 729 bp in length  
\* 41720 41819: gap of 100 bp  
\* 41820 42554: contig of 735 bp in length  
\* 42555 42654: gap of 100 bp  
\* 42655 43361: contig of 707 bp in length  
\* 43362 43461: gap of 100 bp  
\* 43462 44203: contig of 742 bp in length  
\* 44204 44303: gap of 100 bp  
\* 44304 45057: contig of 754 bp in length  
\* 45058 45157: gap of 100 bp  
\* 45158 45898: contig of 741 bp in length  
\* 45899 45998: gap of 100 bp  
\* 45999 46762: contig of 764 bp in length  
\* 46763 46862: gap of 100 bp  
\* 46863 47626: contig of 764 bp in length  
\* 47627 47726: gap of 100 bp  
\* 47727 48445: contig of 719 bp in length  
\* 48446 48545: gap of 100 bp  
\* 48546 49261: contig of 716 bp in length  
\* 49262 49361: gap of 100 bp  
\* 49362 50115: contig of 754 bp in length  
\* 50116 50215: gap of 100 bp  
\* 50216 50931: contig of 716 bp in length  
\* 50932 51031: gap of 100 bp  
\* 51032 51765: contig of 734 bp in length  
\* 51766 51865: gap of 100 bp  
\* 51866 52589: contig of 724 bp in length  
\* 52590 52689: gap of 100 bp  
\* 52690 53433: contig of 744 bp in length

```
* 53434 53533: gap of 100 bp
* 53534 54285: contig of 752 bp in length
* 54286 54385: gap of 100 bp
* 54386 55142: contig of 757 bp in length
* 55143 55242: gap of 100 bp
* 55243 55990: contig of 748 bp in length
* 55991 56090: gap of 100 bp
* 56091 56846: contig of 756 bp in length
* 56847 56946: gap of 100 bp
* 56947 57705: contig of 759 bp in length
* 57706 57805: gap of 100 bp

Query Match      28.5%; Score 535.8; DB 2; Length 76186;
Best Local Similarity 88.7%; Pred. No. 2e-80;
Matches 636; Conservative 0; Mismatches 74; Indels 7; Gaps 5;

QY 492 GTACTTGATGTAATAATTTGTGGGATGATTCCTCTAACT-ATAGTTATCCTTCGTCT 550
   || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 70265 GAACNCGATGTCATAATTTGTGGGATGATTCCTCTAACTNATAGTGCCTTCGTCT 70206

QY 551 -TTCACATGGAGAGTAACACATGCTCTCAAT-TGCGTGTTCATGAGGCTTGTGTATCCTTT 608
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 70205 CNTCACTGGAGAGTAACACATGCTCTCAATTCGTGTTCATGAGGCTTGTGTATCCTTT 70146

QY 609 GCAAGTCTCTCAACAGCAATCAAGTCTTTCGTATCACTTTGGACATATGACATCTCT 668
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 70145 GCAAGTCTTTCGACAGCAATCAAGTCTTTCGTATCACTTTGGACATATGACATCTCT 70087

QY 669 GTAAACCTGCAACCGCAATTCGACATGGGACAGCTGTAATGTTAATGATATCCATT 728
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 70086 GTAAACCTGCAACAGCAATTCGACATGGGACAGCTGTAATGTTAATGATATCCATT 70027

QY 729 TGGATTTTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 788
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 70026 TGGATTTTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 69967

QY 789 CAAAGTGGAAATACCTGGGAAACCAACACACTTTTATGTGTCAGTACAAATGAATAC 848
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69966 CAAAGTGGAAATACCTGGGAAACCAACACACTGCTGTGTCAGTACAAATGAATAC 69907

QY 849 ACTGAACCTGGGAATGATTAATCACTCTTATGACAGATCCCAATTAATTTTTCAC 908
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69906 ACTGAGCTCGGAGTACTATCACTCTTTTGTGTCAGATCCCAATTAATTTTTCAC 69847

QY 909 GTAGTAATGTTAATCAATACACCAAAATACTTCAGGCTCTTAATATTCGAATAGG 968
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69846 ATAGTCAATGTTATCAATACACCAAAATACTTCAGGCTCTTAATATTCGAATAGG 69787

QY 969 AGATTTTCAACAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1028
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69786 AGATTTTCAACAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69728

QY 1029 CAACATGAGGCTACAGACATGTCACAAAGCAGTGGTGGGAGAAATGTAGTCTTTGG 1088
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69727 --ACATGAGACACAGACATGTCACAAAGCAGTGGTGGGAGAAATGTAGTCTTTGG 69670

QY 1089 AGAATTTCAATGTTGTAATTAATTTGCCCTCGGAGAGTGTGAACGACACCGTGA 1148
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69669 AGAATTTCAATGTTGTAATTAATTTGCCCTCGGAGAGTGTGAACGACACCGTGA 69610

QY 1149 CGAAGAACACAAAGAGAGCTTTCAGAGTCTTTATGATTAATTTCTACATTTCTT 1205
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 69609 CGAAGAACACAAAGAGAGCTTTCAGAGTCTTTATGATTAATTTCTACATTTCTT 69553

RESULT 7
AC116548/c
LOCUS
DEFINITION
  Dictyostellium discoideum chromosome 2 map 2886184-2983865 strain
  AX4, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION
  AC116548
VERSION
  AC116548.1 GI:19807722
KEYWORDS
  HTG; HTGS_PHASE2.
```

```
SOURCE
ORGANISM Dictyostellium discoideum.
REFERENCE Dictyostellium discoideum.
AUTHORS 1 (bases 1 to 97683)
          Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
          Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
          Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
          Noegel,A.A.
TITLE Sequence and Analysis of Chromosome 2 of Dictyostellium
JOURNAL Unpublished
REMARK The Dictyostellium Genome Sequencing Consortium
REFERENCE 2 (bases 1 to 97683)
AUTHORS Baumgart,C.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2002) Genome Analysis, Institute of Molecular
          Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT CDS predictions from GeneID may contain errors. Further Information
          is available from IMB Jena, Department of Genome Analysis
          (http://genome.imb-jena.de/dictyostellium/)
          and the University Cologne, Institute for Biochemistry I
          (http://www.uni-koeln.de/dictyostellium/project.shtml)
          Funding
          Agency : Deutsche Forschungsgemeinschaft (DFG).
          * NOTE: This is a 'working draft' sequence.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.

FEATURES
             source
             1..97683
             /organism="Dictyostellium discoideum"
             /strain="AX4"
             /db_xref="taxon:44689"
             /chromosome="2"
             /map="2886184-2983865"
             complement(103..1243)
             /note="ORF_ID:dd_01895"
             /pseudo
             /codon_start=1
             complement(join(2308..2721,2817..3046,3148..3526,3594))
             /note="ORF_ID:dd_01893"
             /pseudo
             complement(4441..4788)
             /note="ORF_ID:dd_01891"
             /codon_start=1
             /product="hypothetical protein"
             /protein_id="AAL99295.1"
             /db_xref="GI:19807723"
             /translation="MPHTSVKTKTNTNQOKYIENWEGYEHVKELIKLRMKYKCSSN
             SLKRICGGEELIHTLSAMNGNTYKIKHVDPSTPVVYNLKRHNLSAYITRKNFS
             PFILVEKTKDDLFM"
             complement(join(5076..6027,6292..6302))
             /note="ORF_ID:dd_01888"
             /pseudo
             /codon_start=1
             complement(join(7498..17492,17678..17936))
             /note="ORF_ID:dd_03234"
             /codon_start=1
             /product="AT1-46.11/98"
             /protein_id="AAL99296.1"
             /db_xref="GI:19807724"
             /translation="MKDIQHLINNGNSPKRVOKLASHIPSTIVOKYIMKLPQDERAP
             DISISMPVICIKFYSIAGSYCKQSNLYKFLBESFRIIAEAGDVLKRNFGNEILAV
             WYNSDKESGEQIFSLRRATRCQAQIRAMFYLISSKEENGKTELPVNSNNINNNIN
             INNNNNNNNNNSGNSFNINNNNTNEOSYSSNMPITPSSSSNNINIOHSDYS
             FTPOFNMTPTSSPIAKSNPSRERVEIEOQKREKREKLINTSSNKERFKOQNVY
             QTSPPSPSFQSQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO
             SFSSTNSGININISNNQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQOQQO
             LTGDGFLGMYNMNRNIEIASGIFLKLQLSKLSKUSKDFSKLSVSVSNVDNNNS
             NNNNNNTVPTSPSVNFSIGSVSNLLGYNKONKONKDKDKSDNNNNNNNNNN
             NVKNHLFSQIIVNRCVLSLKHSCQEVFSDSLFSODNNNNNNNNNNNNNNNNNN
             GTKIETTEITATATTSDDELLGFKVSNIDYFVIFKGMHDVSKVPFPMDKTLEELS
             PEILIQSLLMCHIPHIRVLVTDSTSEYIFLEGHKKLKDYLKNNQSNLSLSNISI
```













Matches	260:	Conservative	0:	Mismatches	291:	Indels	1:	Gaps	1:
Qy 1331	ATTATATGCATTCAC	TAGACAAAAAATTT	CAAAAGGTCTTG	AAAAAGTAAAA	AGTAAAAAGCG		1390		
Db 214154	AAATATATAAAAAAT	ATATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT		214095		
Qy 1391	AGTTGTTCTATAGT	AGAGCGTGATCCCT	CGCCCTAAATA	ATGCTGTAAT	PACACAACTCTTG		1450		
Db 214094	ATATATATATAAAAA	TATATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT		214035		
Qy 1451	GATAGATCCCCAAAA	CAAAAAATTTACCT	TTTGAAGATAGT	CGAAATAGAG	AAAAAAGCG		1510		
Db 214034	TATATATATAAAAA	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT		213975		
Qy 1511	TTTAGTGCCTCAGGT	GTGCACAGACTAG	AGAAAAAGTCTC	AGTTTCACCA	AAATCCACATTC		1570		
Db 213974	ATATATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT		213915		
Qy 1571	AAATGACGTTTTAAAT	TTTAAATTTGTA	AAAAAAGTATTT	TACTGCCCA	AAATATAGA		1630		
Db 213914	TATATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT		213855		
Qy 1631	TTTAAAGTATGGTAT	GTGTTCAATTTCA	ATGTAAGTCT	TAATTAAGAT	FAGGTCATAT		1699		
Db 213854	AAATATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT		213795		
Qy 1690	ATATCAATTTCTCAT	TACCTAAATGAT	TTTGGTCGAT	GGCAGGTTG	TTAAGTACTATC		1749		
Db 213794	ATAAAATATATAAAA	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT		213735		
Qy 1750	ATGTCGTATATTTGT	CCAATATTTATG	TCCACAGAAAA	TATTTCA	TGTAAGTCATATTTT		1809		
Db 213734	ATATATATAATATAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT		213675		
Qy 1810	AAGGAATAAATACAT	AGGCTTAAAAAC	AGTGATTAAC	TTTAAATG	TAAAAAAGAAAAA		1869		
Db 213674	ATATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT	TATATAAAAAAT		213615		
Qy 1870	AAAAAAGAAAAA		1881						
Db 213614	AAATATATATAA		213603						

```

RESULT 12
AL807813/c
LOCUS
DEFINITION
        AL807813          106958 bp    DNA          linear    PRI 26-JUL-2002
        Human DNA sequence from clone Rpl3-487c10 on chromosome X, complete
        sequence.
ACCESSION
        AL807813
VERSION
        AL807813.6      GI:22002803
KEYWORDS
        HTG.
SOURCE
        Homo sapiens.
ORGANISM
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
        Whitehead, S.
TITLE
        Direct Submission
JOURNAL
        Submitted (26-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
        Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
        humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
        On Jul 29, 2002 this sequence version replaced gi:21998278.
COMMENT
        ----- Genome Center
        Center: Wellcome Trust Sanger Institute
        Center code: SC
        web site: http://www.sanger.ac.uk
        Contact: humquery@sanger.ac.uk
        -----

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/Hgp/ChrX> RP13-487C10 is from the library RPC1-13.3 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBACE3.6.

FEATURES	SOURCE
1. <i>Age</i>	1. <i>Age</i>
2. <i>Gender</i>	2. <i>Gender</i>
3. <i>Marital Status</i>	3. <i>Marital Status</i>
4. <i>Education</i>	4. <i>Education</i>
5. <i>Income</i>	5. <i>Income</i>
6. <i>Occupation</i>	6. <i>Occupation</i>
7. <i>Religion</i>	7. <i>Religion</i>
8. <i>Political Affiliation</i>	8. <i>Political Affiliation</i>
9. <i>Health Status</i>	9. <i>Health Status</i>
10. <i>Family Size</i>	10. <i>Family Size</i>
11. <i>Home Ownership</i>	11. <i>Home Ownership</i>
12. <i>Travel Frequency</i>	12. <i>Travel Frequency</i>
13. <i>Employment Status</i>	13. <i>Employment Status</i>
14. <i>Volunteer Work</i>	14. <i>Volunteer Work</i>
15. <i>Charitable Donations</i>	15. <i>Charitable Donations</i>
16. <i>Political Participation</i>	16. <i>Political Participation</i>
17. <i>Community Involvement</i>	17. <i>Community Involvement</i>
18. <i>Life Satisfaction</i>	18. <i>Life Satisfaction</i>
19. <i>Stress Levels</i>	19. <i>Stress Levels</i>
20. <i>Overall Well-being</i>	20. <i>Overall Well-being</i>

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP13-487C10"
/clone_lib="RPCI-13.3"
36290 a 20496 c 20177 g 29995 t
BASE COUNT
ORIGIN

```

[illegible]

RESULT 13

AC007158	AC007158	201299 bp	DNA	linear	PRI 17-AUG-2001
LOCUS	AC007158	201299 bp	DNA	linear	PRI 17-AUG-2001
DEFINITION	Homo sapiens, clone RP11-90A1, complete sequence.				
ACCESSION	AC007158				
VERSION	AC007158.10	GI:11610938			
KEYWORDS	HTG.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 201299)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.				
TITLE	Homo sapiens, clone RP11-90A1				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 201299)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.				

Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,  
 Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,  
 Cooke, P., Dearellano, K., Depayre, E., Devon, K., Dewar, K.,  
 Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,  
 Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,  
 Hagos, B., Heaford, A., Horton, L., Howland, J.C., Jones, C., Kann, L.,  
 Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,  
 Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,  
 Melldrim, J., Mollia, M., Morris, W., Morrow, J., Mychaleckyj, J.,  
 Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,  
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,  
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,  
 Testaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,  
 Wheeler, J., Wu, X., Wyman, D., Ye, W.J. and Zody, M.

## TITLE

## JOURNAL

Submitted (24-MAR-1999) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

## REFERENCE

3 (bases 1 to 201299)

## AUTHORS

Birren, B., Linton, L., Nussbaum, C., Lander, E., Allen, N., Anderson, S.,  
 Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A.,  
 Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,  
 Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,  
 Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,  
 Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,  
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,  
 McPheeters, R., Melldrim, J., Meneus, L., Mihova, T., Mlenga, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Plerre, N., Pollara, V., Raymond, C., Retta, R.,  
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rossetti, M.,  
 Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,  
 Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,  
 Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,  
 Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,  
 Zembek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

Submitted (12-DEC-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Dec 9, 2000 this sequence version replaced gi:11276252.  
 All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L659

Center clone name: 90\_A1

## FEATURES

## source

Location/Qualifiers

1..201299  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="RP11-90A1"  
 /clone\_lib="RP11-11 Human Male BAC"  
 1664..1684  
 /rpt\_family="AT-rich"  
 3468..3598  
 /rpt\_family="L2"  
 3700..3758  
 /rpt\_family="(CAA)n"  
 3785..3839  
 /rpt\_family="L2"  
 complement(4495..5019)  
 /rpt\_family="L1MA9"  
 5029..5080  
 /rpt\_family="(TA)n"

5138..5268  
 /rpt\_family="(TA)n"  
 complement(5271..5356)  
 /rpt\_family="L1MA9"  
 5492..5518  
 /rpt\_family="AT-rich"  
 complement(5520..5825)  
 /rpt\_family="AluJo"  
 complement(5903..6084)  
 /rpt\_family="MLT1A2"  
 6110..6151  
 /rpt\_family="GA-rich"  
 complement(6168..6322)  
 /rpt\_family="MLT1A2"  
 complement(6329..6752)  
 /rpt\_family="MLT1A2-int"  
 complement(6898..6959)  
 /rpt\_family="MLT1A2-int"  
 complement(7018..7078)  
 /rpt\_family="MLT1A2-int"  
 complement(7107..7761)  
 /rpt\_family="MLT1A2-int"  
 complement(7764..8176)  
 /rpt\_family="MLT1A2"  
 8650..8774  
 /rpt\_family="MIR"  
 9716..9858  
 /rpt\_family="CT-rich"  
 10648..11054  
 /rpt\_family="MLT1B"  
 complement(11497..11567)  
 /rpt\_family="MIR"  
 13140..13163  
 /rpt\_family="AT-rich"  
 complement(14648..14845)  
 /rpt\_family="MIR"  
 15776..16075  
 /rpt\_family="AluSx"  
 17569..17855  
 /rpt\_family="AluSg"  
 17856..17878  
 /rpt\_family="(CCAA)n"  
 complement(18262..18315)  
 /rpt\_family="L2"  
 18356..18569  
 /rpt\_family="MERS8A"  
 19351..19399  
 /rpt\_family="(TG)n"  
 20209..20524  
 /rpt\_family="AluYa5"  
 21099..21240  
 /rpt\_family="MER99"  
 21414..21910  
 /rpt\_family="PTRS"  
 21909..21951  
 /rpt\_family="PTRS"  
 complement(22254..22465)  
 /rpt\_family="MER63A"  
 complement(24968..25003)  
 /rpt\_family="MER117"  
 25617..27797  
 /rpt\_family="L1PA7"  
 29160..29208  
 /rpt\_family="(TAAA)n"  
 31294..31333  
 /rpt\_family="(T)n"  
 34460..34497  
 /rpt\_family="A-rich"  
 complement(35336..35535)  
 /rpt\_family="MIR"  
 35562..35657  
 /rpt\_family="GA-rich"  
 35670..35751

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

```
repeat_region      /rpt_family="GA-rich"
35998. .36084
/rpt_family="MIR"
complement(36446. .36608)
/rpt_family="FRAM"
complement(36644. .36776)
/rpt_family="MIR"
complement(37640. .37803)
/rpt_family="MIR"
complement(39017. .39404)
/rpt_family="MLT1A2"
40700. .41014
/rpt_family="(TA)n"
40882
unsure
repeat_region      41018. .41201
/rpt_family="(TATAA)n"
41207. .41268
/rpt_family="(TA)n"
complement(41650. .41786)
/rpt_family="MER58B"
complement(41905. .48040)
/rpt_family="L1PA4"
49825. .49872
/rpt_family="AT-rich"
complement(49888. .50121)
/rpt_family="MLT1D"
50167. .50225
/rpt_family="(TC)n"
complement(50260. .50435)
/rpt_family="MLT1D"
complement(50717. .51018)
/rpt_family="AluSx"
51019. .51159
/rpt_family="MIR"
51466. .51628
/rpt_family="MER104"
51800. .51973
/rpt_family="MIR"
complement(52323. .52794)
repeat_region

Query Match          3.9%; Score 73; DB 9; Length 201299;
Best Local Similarity 52.8%; Pred. No. 0.0029;
Matches 180; Conservative 0; Mismatches 160; Indels 1; Gaps 1;

QY 1542 AAGGTCCTCAGTTCCACCAATCCACATGAGTTTAAATTAATGTAATAACT 1601
Db 40749 ATAAATATATTTTATAATATATATATATATATATATATATATATATAT 40808

QY 1602 GATATTACTGCCAATATATAGAAATATTTTAACTATTTGGTTATGTTAAATTTTCAA 1661
Db 40809 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 40868

QY 1662 TGTGAAATGCTAATPAGATGTCATATATATTCATTTCTTCATCTACTTAATGTTATGG 1721
Db 40859 TATATTTTATATATATATATATATATATATATATATATATATATATATATATATATA 40928

QY 1722 TTGCATGCGAGTTGTTTAAAGTA -CTATCATGTGTATATTTTGTCAATATTTATGTCACAC 1780
Db 40929 TAATATATATTTTATATATATATATATATATATATATATATATATATATATATATAT 40988

QY 1781 AGAAATATTCATGTAAGTCATATATTTTAAAGGAATAATACATAGCCCTTAAACAGTGT 1840
Db 40989 ATGTATATATATATATATATATATATATATATATATATATATATATATATATATATAT 41048

QY 1841 ATAAGCTTTAAATGTAATAAAAAAAAAAAAAAAAAAAAAAAAAA 1891
Db 41049 ATATTTTATATATATATATATATATATATATATATATATATATATATATATATATA 41089
```

```
RESULT 14
AC010903
LOCUS AC010903 91723 bp DNA linear PRI 07-NOV-2001
DEFINITION Homo sapiens BAC clone RP11-550H9 from 2, complete sequence.
```

```
ACCESSION AC010903
VERSION AC010903.8 GI:13162555
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 91723)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 91723)
AUTHORS Mulvaney,E., Haakenson,W. and Belter,E.
TITLE The sequence of Homo sapiens BAC clone RP11-550H9
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 91723)
AUTHORS Waterston,R.H.
TITLE Direct Submision
JOURNAL Submitted (25-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 91723)
AUTHORS Waterston,R.H.
TITLE Direct Submision
JOURNAL Submitted (28-FEB-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 91723)
AUTHORS Waterston,R.
TITLE Direct Submision
JOURNAL Submitted (09-AUG-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 91723)
AUTHORS Waterston,R.
TITLE Direct Submision
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 28, 2001 this sequence version replaced gi:11465184.
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
-----
Summary Statistics
-----
Center project name: H_NH0550H09
-----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RPC1-11 human BAC library was made from the blood of one male donor, as described by Osogawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatenno,M., Cataneese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome







QY	1423	CTAATAATGCTGTTAATACACAACTCTTGGATAGATCCCAAAAGAAACAAAAAATTACCT	1482
Db	9184	TTTGAATTTGAGCTCATACGAGCTAAATGTGATATAGTAATACTATGGATCTAATAAAAAG	9125
QY	1483	TTGAAGATAGTGAATAAGAGAAAAACGTTTAGTGCCTCAGTTGTCACAGACTAGAGAA	1542
Db	9124	TTTAGAGTGGCTTTTATAAGCTTTTTTTTGGCTGGAAATAAAAAAAAATGAAAAAA	9065
QY	1543	AAGTCTCAGTTTACCACCAATCCACATTCAGTTTAAATTTAAATTTGAAAAACTG	1602
Db	9064	AAATTTTTTTTTAAAAATAAAAAAAGAAAGAAAAATAAAAAAATTTAAAAAAA	9005
QY	1603	ATATTACTGCCAAATATAAGAAAAATATTTAAGTATTTGGTTATGTGTAATTTTCAAT	1662
Db	9004	AAATAAAAAAAATAAAATAAAAAAAATAAAAAAAAATAAAAAACATAAAT	8945
QY	1663	GTGAAATGCTAATTAGATAGTGCATATATATCAATTTCTCAATCTTAATGATTTGT	1723
Db	8944	CAAAATTTGGACGCTTGGGGAACATTATAAACTAAAAAGTTTTTTTTTTTTTTTCT	8885
QY	1723	TGCATGGCAGTTTGTTAAAGTACATACATGTTGTAATTTT--TGCAATATTTATGTCAC	1780
Db	8884	TATATTTTTCATGATTTAAATATATAAAAAATATATTTTTTTTAAATATATATATTTAAT	8825
QY	1781	AGAAAAATTCATCTAGTGCATATTTTTTTAGGAATAAATACATAGCCTTAAAAACATGT	1840
Db	8824	TATTTTTAATTTATTTTAAATTTTTTATTTTATTTTATTTTGGTTTTTCCCAAAATAA	8765
QY	1841	ATAACTTTAAATGTAAAAAATAAAAAAATAAAAAAATAAAAAA	1881
Db	8764	ATAAATTTAAAAAATAACATAAAAAATAAAAAATAAAAAACATAAA	8724

Search completed: December 10, 2002, 09:25:26  
Job time : 6021 secs

<p> <code>/note="ORF_ID:dd_01625"</code>  <code>/codon_start=1</code>  <code>/product="hypothetical protein"</code>  <code>/protein_id="AAL93606.1"</code>  <code>/db_xref="GI:19703144"</code>  <code>/translation="MKLLIIVLFLKLSINGFENIDGNLLETIVETISFNGFNES</code>  <code>NKFDKINGDSSICFSGVGGGFNVLFSSGLGNSEIFVKSSEDFYFEDHYETCKKHQ</code>  <code>IGITDLHVPKNLKRNTFFVQGEIYISGCFNIYLNKCDIKTDKNDYSIYEQIL</code>  <code>KIHSLPASHFTSPFHFIFTFITFLVFTFTNSLSFLLIINISQGLTIFGFR</code>  <code>FAIIPNCAIFSTETIGLSAYLVESNFGYGEYRNPFKSNVGYVLAIQGLTIFGI</code>  <code>FENKFIISLIDYQQQSLQVLRKFIISIIWISIPISIIISIFIVDNNRFSLSIIN</code>  <code>DGFNIFYLLIIVYHPCKSNILVKRLNSYYDDGDELSINQNSQQKSIDENLENLY</code>  <code>NK"</code>  <code>complement(join(17079..19519,19709..20690))</code>  <code>/note="ORF_ID:dd_02339"</code>  <code>/codon_start=1</code>  <code>/product="hypothetical protein"</code>  <code>/protein_id="AAL93607.1"</code>  <code>/db_xref="GI:19703145"</code>  <code>/translation="MKTSHKKKSLGGLFSHSSSPNLKSFTEEVIHQOQOQOQHNK</code>  <code>NNNNNNHQRQPTSTSSYIDSSASSIEPTSGYLSKTSSSSLSPSFLHNQNGNQ</code>  <code>NQLNSSGSLNNQOQOQOQISQTSISPTSPMGKKKFRSLAQAFKPKVKNNGNK</code>  <code>LKKEELTKNGLNQOQTHSNLFPNVVSEIDINGSGGVTSTGTNIISHKSPSSSS</code>  <code>SSSSKHKQRKRFETISQSTEDYDIPRIKMSVEYLFKCLVPGIFRESANME</code>  <code>LARLSQLEKQGDIDLSYTOPHCIGGLKLYIFREKPIFPYDLHKRIYSVLNEEDS</code>  <code>NKLVKSLGSLGKGQGLILRYLFELLNAVSTNCNMFMYONLAICFAPSLIQSFDL</code>  <code>SCYDLIERNIENHIFGVVVPVVDNSNVSSSILAHPIIGKSETNVSLNLIATV.LI</code>  <code>DGIDLNNSSNNSSNNSSSPYKPKPKPKPKPKLNRRNSISPKLSSSTSRKSI</code>  <code>NLSNMSILSDVQEQEQOQOQDEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ</code>  <code>RPKGTVOYLNRINCRPSSVNNNRNIKQOHHHHHHQOQOQOHHQOQOQSSSES</code>  <code>NSSLTSSQPKRSLVSEEGKNGVNEIYSLKEESKLKPKSLKSLRQMTIYVNSN</code>  <code>IGNGEKDCDCTDEBELKISKPIIRSTVNEILKETDNDNNNDQIINNNSNNI</code>  <code>TKTTTTNTTTNTTNTSISQSIITSAKPTPTPAPMOTSFSLSTKOTNSP</code>  <code>SSSPSTSVSTSSSSLSISDNKTMNVNVPQAPNVPSSPNVFNFSVPT</code>  <code>TTTTSIGNSVFSYFRIKRTTPSIFSPKOSICKPKNTSSLSSSSNFSKSTNP</code>  <code>TPAPALITSLVNSTSSSKLQKALPTTFTTQTSPLLILOPLTSTSTSTSTS</code>  <code>TSTSTSTSTSTTTPITPTTSSFKNSFSTSDNNSNNNNNNNNNNNNNNRNTAP</code>  <code>NNNNVTSPITVSSPSTPSITPSSTTFIKGHSNLSLSQTPSSSSSSSLGNRLTAP</code>  <code>RLQNLHTSSLSIKFNIDTPSGINQMGKGVVQKQDPVKSLAIFERE"</code>  <code>complement(join(22671..22910,23013..23359,23480..23555))</code>  <code>/note="ORF_ID:dd_02340"</code>  <code>/codon_start=1</code>  <code>/product="hypothetical protein"</code>  <code>/protein_id="AAL93608.1"</code>  <code>/db_xref="GI:19703146"</code>  <code>/translation="MVKILPLVSLVILVSACFTIGFTSYQVKYEGVSGSTVTTY</code>  <code>YKYNKVSLENGEEDTSDYDDGEFSAQLFKYKALSMDLIAFVNLFSSACMY</code>  <code>GAFAFKETSPANOTKLKSLNGATVLIATFVVLGPKASRECEDEYAKDSSLNLYG</code>  <code>SOVDEMCGDLYKKFISGTDDEGSKLSWGPTTAWIAIVVGLGIAFFNIPIYCGPVP</code>  <code>RS"</code>  <code>24236..25094</code>  <code>/note="ORF_ID:dd_02341"</code>  <code>/codon_start=1</code>  <code>/pseudo</code> </p>	<p> CDS  10457 a 2488 c 2647 g 9525 t </p>
<p> CDS  BASE COUNT  ORIGIN </p>	

	Query Match	3.8%;	Score 71.8;	DB 2;	Length 25117;
	Best Local Similarity	45.6%;	Pred. No. 0.0074;		
	Matches 292;	Conservative 0;	Mismatches 347;	Indels 2;	Gaps 1;
QY	1243	TTTTATGTTTTAGGCCCAAGTCACCTTTTCTAGTAAANAATTAAGATTGCTGTTTTTTTACTCATGG	1302		
Db	9364	TCATATTATTATATATGTFATATTGTTTTTAATATTCTCTTAATGGTTTTTTTTTTTTTTTT	9305		
QY	1303	CTTATGGAAACAACATATATTTTCACCCCTCTATTATATGTCATTCACTACACAAAAATTTCAAA	1362		
Db	9304	TTTTCTTTCATATTTTAAACTTTATTTCTCAAATAATATCGCGAGAAAAAAAATAAAA	9245		
QY	1363	AGGCTCTGAAAGCTAAATGAAAGCGGAGTCTGTTCTTATAGTAGAAGCTGATCCCGTCG	1422		
Db	9244	TAATAATAATTAATAATTAATAGATATAGCATTTATTTATTTACTTACACAGCTGAGTTAAGAT	9185		

**This Page Blank (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 10, 2002, 07:33:36 ; Search time 398 Seconds

(without alignments)  
10643.238 Million cell updates/sec

Title: US-09-845-721-1

Perfect score: 1881

Sequence: 1 gttattttcctaaaggaaa.....aaaaaaaaaaaaaaaaaaaaa 1881

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB-seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N\_Geneseq\_101002.\*
- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
  - 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
  - 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
  - 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
  - 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
  - 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
  - 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
  - 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
  - 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
  - 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
  - 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
  - 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
  - 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
  - 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
  - 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
  - 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
  - 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
  - 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
  - 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
  - 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
  - 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
  - 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
  - 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
  - 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1881	100.0	1881	24	Human GPR22 encodi
2	1879.4	99.9	1881	24	Cysteine protease-
3	1625	86.4	2238	20	Human G-protein co
4	1302	69.2	1302	21	Human G protein-co
5	1297.2	69.0	1302	21	DNA encoding human
6	170.2	9.0	200	24	Cysteine protease-
7	157	8.3	200	24	Cysteine protease-
8	67.6	3.6	5487	24	Human immune syste
9	65.8	3.5	7857	24	Human angiogenesis

C 10	63.8	3.4	8842	24	ABL33448	Human immune syste
C 11	63.8	3.4	8842	24	AAS63334	Chemically pretrea
C 12	63.4	3.4	61020	22	AAS46788	Tumour suppressor
C 13	63	3.3	11555	24	ABL32616	Human immune syste
C 14	63	3.3	11555	24	AAD28380	Human chemically t
C 15	62.8	3.3	17280	22	AAS46772	Tumour suppressor
C 16	62.4	3.3	50000	22	ABL55643	AMEPV genome fragm
C 17	61.8	3.3	5255	24	ABX28419	DNA transcription
C 18	61.6	3.3	18683	24	ABL54334	Chemically treated
C 19	61.6	3.3	18683	24	ABL32313	Human immune syste
C 20	61.4	3.3	18988	22	AAS46342	Tumour suppressor
C 21	61.4	3.3	18988	24	ABL70204	Chemically treated
C 22	61.4	3.3	18988	24	ABL32701	Human immune syste
C 23	61.4	3.3	18988	24	ABL34509	Human metastasis a
C 24	61.2	3.3	2104	13	AAQ25273	Sequence encoding
C 25	61.2	3.3	6050	24	ABL34011	Human immune syste
C 26	61	3.2	7319	24	ABL34045	Human immune syste
C 27	60.8	3.2	5487	24	ABL33599	Human immune syste
C 28	60.8	3.2	6052	24	AAS61094	Human gene regulat
C 29	60.6	3.2	5276	24	ABL32150	Human immune syste
C 30	60.4	3.2	545	22	AAH70126	Human cervical can
C 31	60.4	3.2	5447	22	AAS46758	Tumour suppressor
C 32	60.2	3.2	5852	12	AAQ11710	Dictyostellum plas
C 33	60.2	3.2	10048	24	ABL70314	Chemically treated
C 34	60.2	3.2	10048	24	AAS61252	Human gene regulat
C 35	60	3.2	60	24	ABN45479	Human spliced tran
C 36	60	3.2	621	22	AAH71551	Human cervical can
C 37	60	3.2	883	22	ABL15210	Human breast cance
C 38	59.8	3.2	3873	24	ABO67089	Human angiogenesis
C 39	59.8	3.2	5675	24	ABL33276	Human immune syste
C 40	59.6	3.2	6035	24	ABL33525	Human immune syste
C 41	59.6	3.2	11790	24	ABL32543	Human immune syste
C 42	59.4	3.2	10682	24	ABN80085	Human chemically m
C 43	59.2	3.1	6242	24	ABL34148	Human immune syste
C 44	59	3.1	5487	24	ABL33599	Human immune syste
C 45	59	3.1	6292	22	AAS46735	Tumour suppressor

ALIGNMENTS

RESULT 1

ABA05328

ID ABA05328 standard; cDNA; 1881 BP.

AC ABA05328;

DT 27-FEB-2002 (first entry)

XX Human GPR22 encoding cDNA.

DE Human GPR22 encoding cDNA.

XX Human; GPR22; anorectic; appetite control agent; GPR; gene;

KW G protein-coupled receptor; orphan receptor; antisense gene therapy;

KW mouse; GPR56; SS.

XX Homo sapiens.

OS Homo sapiens.

XX Key

XX CDS

XX Location/Qualifiers

XX 237..1538

XX /\*tag=

XX /product= "GPR22"

XX

XX WO200183550-A2.

XX

XX 08-NOV-2001.

XX

XX 30-APR-2001; 2001WO-GB01874.

XX

XX 03-MAY-2000; 2000US-201418P.

XX (ASTR ) ASTRAZENECA AB.

XX (ASTR ) ASTRAZENECA UK LTD.

XX

XX

PI Brenand JC, Hart KA;  
XX WPI; 2002-066519/09.  
DR P-PSDB; AAM47903.  
XX  
PT Identifying appetite control agent for controlling obesity, comprises  
PT screening agonists or antagonists of G protein-coupled receptor, GPR22,  
PT and using them as test compounds in appetite control test procedures -  
XX  
PS Disclosure; Page 14-15; 21pp; English.  
XX  
CC The invention relates to identifying an anorectic appetite control agent,  
CC comprising screening for agonists and/or antagonists of G  
CC protein-coupled receptor GPR22 (an orphan receptor), using one or more  
CC agonists and/or antagonists so identified as test compounds in one or  
CC more appetite control procedures and selecting an active compound for use  
CC as an appetite control agent. An antisense oligonucleotide to the GPR22  
CC gene is also useful for controlling obesity using antisense gene therapy.  
CC Note: The GPR22 encoding cDNA sequence (ABA05328) and encoded protein  
CC (AAM47903) are disclosed, however two DNA sequences described as human  
CC (ABA05329) and mouse (ABA05328) GPR56 are given in the sequence listing  
CC but are not otherwise referred to in the specification.  
XX  
SQ Sequence 1881 BP; 649 A; 317 C; 288 G; 527 T; 0 other;

Query Match 100.0%; Score 1881; DB 24; Length 1881;  
Best Local Similarity 100.0%; Pred. No. 2.2e-314;  
Matches 1881; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATTTCTTCAAAAGGAAACACAAATTTCTTTTATATCAAAACAAATGCAAACTTGATG 60  
DB 1 GTATTTCTTCAAAAGGAAACACAAATTTCTTTTATATCAAAACAAATGCAAACTTGATG 60  
QY 61 GTTCTTAATCTACATTTTCTAATTAATAGTTTACAAACTTAAATTAAGTACAC 120  
DB 61 GTTCTTAATCTACATTTTCTAATTAATAGTTTACAAACTTAAATTAAGTACAC 120  
QY 121 AATTGAAGATTTTCTTTTACAAAGAACACGTTATACGTCATTTAAATGGCCAAATA 180  
DB 121 AATTGAAGATTTTCTTTTACAAAGAACACGTTATACGTCATTTAAATGGCCAAATA 180  
QY 181 TCAATAGTTTATCTATTCTACTTTCTAGGGAAGAAACCAACGTCGCCAAAGAATGT 240  
DB 181 TCAATAGTTTATCTATTCTACTTTCTAGGGAAGAAACCAACGTCGCCAAAGAATGT 240  
QY 241 GTTTTCTCCCATCTCGAAATCAACATGACGCTCTGAATCTAACATTCACAGTCGAGATG 300  
DB 241 GTTTTCTCCCATCTCGAAATCAACATGACGCTCTGAATCTAACATTCACAGTCGAGATG 300  
QY 301 ACATGATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTCAAG 360  
DB 301 ACATGATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTCAAG 360  
QY 361 TGTCTCTACCGGATTTCTTATGTAGAAATGTCTGGGACTTGGCAGCAACCTCACATG 420  
DB 361 TGTCTCTACCGGATTTCTTATGTAGAAATGTCTGGGACTTGGCAGCAACCTCACATG 420  
QY 421 TATTGGTACTTTTACGTGATGAATCCAACTTAATCAACTCTCTCAGTAACTATTACAA 480  
DB 421 TATTGGTACTTTTACGTGATGAATCCAACTTAATCAACTCTCTCAGTAACTATTACAA 480  
QY 481 TGAATCTTCATGACTTGATGPAATAATTTGTGGGATGATTCCTCTAATATAGTTA 540  
DB 481 TGAATCTTCATGACTTGATGPAATAATTTGTGGGATGATTCCTCTAATATAGTTA 540  
QY 541 TCCCTCTGCTTTCATCGAGAGTACACTGCTCTCATTTGCTGTTTCCATGAGGCTTGTG 600  
DB 541 TCCCTCTGCTTTCATCGAGAGTACACTGCTCTCATTTGCTGTTTCCATGAGGCTTGTG 600  
QY 601 TATCTTTTGCAGTCTCTCAACAGCAATCAACGTTTTTGTCTATCACTTTGGACAGATATG 660  
DB 601 TATCTTTTGCAGTCTCTCAACAGCAATCAACGTTTTTGTCTATCACTTTGGACAGATATG 660

QY 661 ACATCTCTGTAAACACCTGCAAAACCGAAATTTCTGACAATGGCAGAGCTGTAAATGTAATGA 720  
DB 661 ACATCTCTGTAAACACCTGCAAAACCGAAATTTCTGACAATGGCAGAGCTGTAAATGTAATGA 720  
QY 721 TATCCATTTGGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 780  
DB 721 TATCCATTTGGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 780  
QY 781 TCAGTCTTCAAAAGTGGAAATACCTGGGAAACAAAGACACTTTTATGTGTGTCAGTACAAATG 840  
DB 781 TCAGTCTTCAAAAGTGGAAATACCTGGGAAACAAAGACACTTTTATGTGTGTCAGTACAAATG 840  
QY 841 AATACTACACTGAACCTGGGAATGTATATACCTGTTAGTACAGATCCCAATATTTCTTTT 900  
DB 841 AATACTACACTGAACCTGGGAATGTATATACCTGTTAGTACAGATCCCAATATTTCTTTT 900  
QY 901 TCACCTGTTGTAGTAAATGTTAATCATACATACACCAAAATCTTCAGGCTCTTAATATTCGAA 960  
DB 901 TCACCTGTTGTAGTAAATGTTAATCATACATACACCAAAATCTTCAGGCTCTTAATATTCGAA 960  
QY 961 TAGGCACAAAGATTTTCAACAGGGCAGAGAAAGAAAGCAAGAAAGAAAGCAAAATTTCTC 1020  
DB 961 TAGGCACAAAGATTTTCAACAGGGCAGAGAAAGAAAGCAAGAAAGAAAGCAAAATTTCTC 1020  
QY 1021 TAACACACAAACATGAGGCTACAGACATGTTCACAAAGCAGTGGTGGGAGAAATGTAGTCT 1080  
DB 1021 TAACACACAAACATGAGGCTACAGACATGTTCACAAAGCAGTGGTGGGAGAAATGTAGTCT 1080  
QY 1081 TTGGTGTAAAGAACTTCAGTTTCTGTAAATTAATTTGCCCTCGGGCAGCTGTGAAACGACACC 1140  
DB 1081 TTGGTGTAAAGAACTTCAGTTTCTGTAAATTAATTTGCCCTCGGGCAGCTGTGAAACGACACC 1140  
QY 1141 GTGAACGACGAGAAAGACAAAAGAGAGTCTTCAGGATGTCCTTTATTTATTTTCTACAT 1200  
DB 1141 GTGAACGACGAGAAAGACAAAAGAGAGTCTTCAGGATGTCCTTTATTTATTTTCTACAT 1200  
QY 1201 TTCTTCTCTGCTGGACACCAATTTCTGTTTTAAATACCACCAATTTATGTTTTAGGCCCAA 1260  
DB 1201 TTCTTCTCTGCTGGACACCAATTTCTGTTTTAAATACCACCAATTTATGTTTTAGGCCCAA 1260  
QY 1261 GTGACCTTTTGTAGTAAATTAAGATTTGTTTTTGTAGTCATGCGCTTATGGAACAACTATAT 1320  
DB 1261 GTGACCTTTTGTAGTAAATTAAGATTTGTTTTTGTAGTCATGCGCTTATGGAACAACTATAT 1320  
QY 1321 TTCACCTCTATTATATGCAATTCAGTACACAAATTTTCAAAAGGCTTTGAAAGTAAATA 1380  
DB 1321 TTCACCTCTATTATATGCAATTCAGTACACAAATTTTCAAAAGGCTTTGAAAGTAAATA 1380  
QY 1381 TGAAGAACGAGTTGTTTCTATATAGTAGAGCTGATCCCTGCTTAATATGCTGTAATAC 1440  
DB 1381 TGAAGAACGAGTTGTTTCTATATAGTAGAGCTGATCCCTGCTTAATATGCTGTAATAC 1440  
QY 1441 ACAACTCTGGATAGATCCCAAGAAACAAACAAATTAACCTTTGAAGATAGTGAATAA 1500  
DB 1441 ACAACTCTGGATAGATCCCAAGAAACAAACAAATTAACCTTTGAAGATAGTGAATAA 1500  
QY 1501 GAGAAAAACGTTTACGTGCTCAGTTGTACAGACTAGAGAAAACTCTCAGTTTCCACCA 1560  
DB 1501 GAGAAAAACGTTTACGTGCTCAGTTGTACAGACTAGAGAAAACTCTCAGTTTCCACCA 1560  
QY 1561 ATCCACATTCAAATGAGTTTTTAAATTTTAAATTTGTAAGTATGCTGCAATATAGT 1620  
DB 1561 ATCCACATTCAAATGAGTTTTTAAATTTTAAATTTGTAAGTATGCTGCAATATAGT 1620  
QY 1621 AGAAAAATATTTTAAAGTATGTTGTTGTAATTTTCAATGCTGAATGCTAATTAGAT 1680  
DB 1621 AGAAAAATATTTTAAAGTATGTTGTTGTAATTTTCAATGCTGAATGCTAATTAGAT 1680  
QY 1681 AGGTCATATATATCAATTTCTTCAATTAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1740  
DB 1681 AGGTCATATATATCAATTTCTTCAATTAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1740  
QY 1741 AGTACTATCATGTGTATATTTTGTCAATATATATGTCACAGAAAAATATTCATGTAAGTC 1800

```
|||||
Db 1741 ATACTATCATGATGATATTTTGTCAATATATGTCACAGAAATATTCATGTAAGTC 1800
QY 1801 ATATTTTAAAGAAATTAATACATAGCCTTAAACAGTGTATTAACCTTAAATGTAATAA 1860
Db 1801 ATATTTTAAAGAAATTAATACATAGCCTTAAACAGTGTATTAACCTTAAATGTAATAA 1860
QY 1861 AAAAAAAAAAAAAAAAAAAAAA 1881
Db 1861 AAAAAAAAAAAAAAAAAAAAAA 1881

RESULT 2
ABA91703
ID ABA91703 standard; cDNA; 1881 BP.
XX
AC ABA91703;
XX
DT 01-MAY-2002 (first entry)
XX
DE Cysteine protease-like gene.
XX
KW Cysteine protease; mouse; transgenic mouse; transgenic animal;
KW animal model; gene disruption; gene targeting; gene detection;
KW gene therapy; enzyme; gene; ss.
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 237..1538
FT /tag= a
FT /product= "Cysteine_protease"
FT misc_feature 1..395
FT /tag= b
FT /note= "sequence flanking Neo insert in targeting
FT construct"
FT misc_feature 396..569
FT /tag= c
FT /note= "deleted in targeting construct"
FT misc_feature 570..1117
FT /tag= d
FT /note= "sequence flanking Neo insert in targeting
FT construct"
XX
PN WO200206445-A2.
XX
PD 24-JAN-2002.
XX
PE 17-JUL-2001; 2001WO-US22402.
XX
PR 19-JUL-2000; 2000US-219168P.
PR 19-JUL-2000; 2000US-219171P.
PR 19-JUL-2000; 2000US-219175P.
PR 25-JUL-2000; 2000US-221455P.
PR 13-DEC-2000; 2000US-256212P.
XX
PA (DELT-) DELTAGEN INC.
XX
PI Brennan TJ, Allen KD;
XX
DR WPI: 2002-179787/23.
DR P-PSDB; AAM50840.
XX
PT Non-human transgenic animal useful as a disease model and for
PT identifying agents that modulate gene expression and function,
PT comprises a disruption in a targeted gene e.g. cysteine protease-like
PT gene
XX
PS Example 1; Fig 1; 6lpp; English.
XX
CC The present sequence is that of full-length cDNA for the mouse
CC cysteine protease-like gene. The invention provides non-human
CC transgenic animals containing targeted gene disruptions, including
```

```
CC gene disruptions of the cysteine protease-like gene. A claimed
CC gene targeting construct comprises a first polynucleotide sequence
CC homologous to a target gene, a second polynucleotide sequence
CC homologous to the target gene, a selectable marker and optionally
CC a screening marker. The targeting construct is obtained by
CC inserting a first and a second polynucleotide homologous to
CC different regions of the target gene into a vector so that they
CC flank a positive selection marker. A cell, especially a murine
CC embryonic stem cell, and a transgenic animal comprising a
CC disruption in a target gene are claimed. In the present case,
CC a cysteine protease-like-specific targeting construct having the
CC ability to disrupt or modify cysteine protease-like genes was
CC created using the targeting arms (homologous sequences) given
CC in ABA91704 and ABA91705. Transgenic mice were generated. The
CC cell- and animal-based systems are useful as models for disease
CC and for identifying agents that modulate gene expression and
CC function, and as potential treatments for various disease states
CC and disease conditions. Methods of treating diseases associated
CC with disrupted targeted gene expression or function comprise
CC detecting and replacing mutated target genes through gene therapy.
XX
SQ Sequence 1881 BP; 649 A; 318 C; 288 G; 626 T; 0 other;

Query Match 99.9%; Score 1879.4; DB 24; Length 1881;
Best Local Similarity 99.9%; Pred. No. 4 4e-314;
Matches 1880; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTTATTCTTCAAAGGAAACACAAATTTCTTTTATATCAAAACAAATGCAAACTTGATG 60
Db 1 GTTATTCTTCAAAGGAAACACAAATTTCTTTTATATCAAAACAAATGCAAACTTGATG 60
QY 61 GTTCTTAATCTACATTTCTATTATAGTTTACAACTTAAANATTAAGTACAC 120
Db 61 GTTCTTAATCTACATTTCTATTATAGTTTACAACTTAAANATTAAGTACAC 120
QY 121 AATTGAAAGATTTTCTTCTTACAAAGAACAGTTATAGTCAATTTAAATTCGAATA 180
Db 121 AATTGAAAGATTTTCTTCTTACAAAGAACAGTTATAGTCAATTTAAATTCGAATA 180
QY 181 TCAATAGTTTATTTTATTTTACCTTTCTAGGGAAAAAACCACACTGCTCCAAAGAAATGT 240
Db 181 TCAATAGTTTATTTTATTTTACCTTTCTAGGGAAAAAACCACACTGCTCCAAAGAAATGT 240
QY 241 GTTTTCTCCCATCTCGGAAATCAACATGAGTCTGAATCTACATTTACAGTGGAGATG 300
Db 241 GTTTTCTCCCATCTCGGAAATCAACATGAGTCTGAATCTACATTTACAGTGGAGATG 300
QY 301 ACATTGATGACATCAACACCAATATGTACCACCACTATCATATCCGTTAAGCTTTCAAG 360
Db 301 ACATTGATGACATCAACACCAATATGTACCACCACTATCATATCCGTTAAGCTTTCAAG 360
QY 361 TGTCTCTCACCGGATTTCTTTATGTTAGAAATTTGTTGGGACTTGGCAGCAACCTCACTG 420
Db 361 TGTCTCTCACCGGATTTCTTTATGTTAGAAATTTGTTGGGACTTGGCAGCAACCTCACTG 420
QY 421 TATTGGTACTTTTACTGCATGAATCCAACTTAATCAACTCTGCAGTAAACATATTACAA 480
Db 421 TATTGGTACTTTTACTGCATGAATCCAACTTAATCAACTCTGCAGTAAACATATTACAA 480
QY 481 TGAATCTTCATGTTGATGTAATATTTGTTGGGATGTTATTCCTCTAATCTATAGTTA 540
Db 481 TGAATCTTCATGTTGATGTAATATTTGTTGGGATGTTATTCCTCTAATCTATAGTTA 540
QY 541 TCCTTCTGCTTTTCACTGGAGAGTAAACACTGCTCTCATTTTCTGTTTCCATGAGGCTTGTG 600
Db 541 TCCTTCTGCTTTTCACTGGAGAGTAAACACTGCTCTCATTTTCTGTTTCCATGAGGCTTGTG 600
QY 601 TATCTTTTGCAGTGTCTCAACAGCAATCAACAGTATTTTGTGTATCACTTTTGACAGATATG 660
Db 601 TATCTTTTGCAGTGTCTCAACAGCAATCAACAGTATTTTGTGTATCACTTTTGACAGATATG 660
QY 661 ACATCTCTGTAAACCTTGCAACCGAATTTCTGACAAATGGGAGAGCTGTATGTTATGA 720
Db 661 ACATCTCTGTAAACCTTGCAACCGAATTTCTGACAAATGGGAGAGCTGTATGTTATGA 720
```



Db	144	AAATCAACATGCGAGTCTGAATCTACATATTACAGTGGGAGATGACATTTGATGACATCAACA	203
Qy	319	CCAATATGTACCAACCACCTATCATATCGGTTAAAGCTTTCAAGTGTCTCTCACGGGATTTTC	378
Db	204	CCAATATGTACCAACCACCTATCATATCCGTTAAAGCTTTCAAGTGTCTCTCACGGGATTTTC	263
Qy	379	TTATGTTAGAAATTTGTTGGGACTTGGCAGCAACCTCACATGTAATTTGGTACTTTTACGCA	438
Db	264	TTATGTTAGAAATTTTTTGTGGACTTTGGCAGCAACCTCACATGTAATTTGGTACTTTTACTGCA	323
Qy	439	TGAATCCCAACTTATCAACTCTGTCAGCTACATATTACAAATCAATCTTCATGTACTCTTG	498
Db	324	TGAATCCCAACTTTAATCAACTCTGTCAGTAACATATTACAAATGAATCTTCATGTACTCTTG	383
Qy	499	ATGTAATAATTTGTTGGGATGATTTCCCTCAACTATAGTTATTCCTCTCTGCTTTCACCTGG	558
Db	384	ATGTAATAATTTGTTGGGATGATTTCCCTCAACTATAGTTATTCCTCTCTGCTTTCACCTGG	443
Qy	559	AGAGTAACACTGCTCTCAATTTGCTCTTTCCATGAGGCTTGTAATCTTTTGCAGTGCTCT	618
Db	444	AGAGTAACACTGCTCTCAATTTGCTCTTTCCATGAGGCTTGTAATCTTTTGCAGTGCTCT	503
Qy	619	CAACAGCAATCAACGTTTTTGTCTATCACTTTTGGCAGATATGACATCTCTGTAAACCTGG	678
Db	504	CAACAGCAATCAACGTTTTTGTCTATCACTTTTGGCAGATATGACATCTCTGTAAACCTGG	563
Qy	679	CAAACCGAATCTTGACAATGGGCAGAGCTGTAATGTTAATGATATCCATTTGGAGTTTTTT	738
Db	564	CAAACCGAATCTGACNATGGGCAGAGCTGTAAATGTTAATGATATCCATTTGGAGTTTTTT	623
Qy	739	CTTTTTTCTCTTTCCTGATTCCTTTTATGAGGTAATAATTTTTTTCAGTCTTCAAGTGGAA	798
Db	624	CTTTTTTCTCTTTCCTGATTCCTTTTATGAGGTAATAATTTTTTTCAGTCTTCAAGTGGAA	683
Qy	799	ATACCTGGGAAACAGACACATTTTATGTGTCAGTACAAATGAATACTACACTGAACCTGG	858
Db	684	ATACCTGGGAAACAGACACATTTTATGTGTCAGTACAAATGAATACTACACTGAACCTGG	743
Qy	859	GAATGTTATATCACCTGTTAGTACAGATCCCAATATCTTTTTTCACCTGTTGTAGTAATGT	918
Db	744	GAATGTTATATCACCTGTTAGTACAGATCCCAATATCTTTTTTCACCTGTTGTAGTAATGT	803
Qy	919	TAATACATATACACAAAATATCTTCAAGCTCTTTAATATTCGAATAGGCACAAGATTTTCAA	978
Db	804	TAATACATATACACAAAATATCTTCAAGCTCTTTAATATTCGAATAGGCACAAGATTTTCAA	863
Qy	979	CAGGGCAGAGAAGAAAGCAAGAAAGAAAGACAAATTTCTTAACACACACACATCAGG	1038
Db	864	CAGGGCAGAGAAGAAAGCAAGAAAGAAAGACAAATTTCTTAACACACACACATCAGG	923
Qy	1039	CTACAGACATGTACAAAGCAGTGTGGGAGAAATGTAGTCTTTGGTGTAGAACTTTCAG	1098
Db	924	CTACAGACATGTACAAAGCAGTGTGGGAGAAATGTAGTCTTTGGTGTAGAACTTTCAG	983
Qy	1099	TTTTCTGTATAATTGCCCCCTCGGCGAGCTGTGAAACGACACCGTGAACGACGAGAAAGAC	1158
Db	984	TTTTCTGTATAATTGCCCCCTCGGCGAGCTGTGAAACGACACCGTGAACGACGAGAAAGAC	1043
Qy	1159	AAAAGAGAGTCTTCAGGATGTCTTTATGATATTATTTCTACATTTCTCTCTGCTGGACAC	1218
Db	1044	TAAAGAGAGTCTTCAGGATGTCTTTATGATATTATTTCTACATTTCTCTCTGCTGGACAC	1103
Qy	1219	CAATTTCTGTTTTTAAATACCACTTTTATGTTTATAGGCCCAAGTGACCTTTTACTTAAAT	1278
Db	1104	CAATTTCTGTTTTTAAATACCACTTTTATGTTTATAGGCCCAAGTGACCTTTTACTTAAAT	1163
Qy	1279	TAAGATTGTGTTTTTTTATGTCATGGCTTATGGAACAACATATATTTTCAACCTCTATTATATG	1338
Db	1164	TAAGATTGTGTTTTTTTATGTCATGGCTTATGGAACAACATATATTTTCAACCTCTATTATATG	1223
Qy	1339	CATTCTACTAGACAAAATTTTCAAAGGCTTGTGAAAGTAAATGAAAGGCGAGTCTGTTT	1398
Db	1224	CATTCTACTAGACAAAATTTTCAAAGGCTTGTGAAAGTAAATGAAAGGCGAGTCTGTTT	1283

Qy	1399	CTATAGTAGAGCTGATCCCGCTCCCTAATAATAGCTGTGTAATACACAACATCTTTGGATAGATC	1458
Db	1284	CTATAGTAGAGCTGATCCCGCTCCCTAATAATAGCTGTGTAATACACAACATCTTTGGATAGATC	1343
Qy	1459	CCAAAGCAACAAAAAATTTACCTTTGAAGATAGTGAATTAAGAGAAAAACGTTTACGTGC	1518
Db	1344	CTAAAGCAACAAAAAATTTACCTTTGAAGATAGTGAATTAAGAGAAAAATGTTTATGTGC	1403
Qy	1519	CTCAGGTTGTCACAGACTAGAGAAAAAGTCTCAGTTTCCACCAATCCCAATTCCAAATGAGT	1578
Db	1404	CTCAGGTTGTCACAGACTAGAGAAAAAGTCTCAGTTTCCACCAATCCCAATTCCAAATGAGT	1463
Qy	1579	TTTAAATTTAAATGTTAAAACTGATATTACTTGCACCAATATAAGAAAAATATTTTAAGTA	1638
Db	1464	TTTAAATTTAAATGTTAAAACTGATATTACTTGCACCAATATAAGAAAAATATTTTAAGTA	1523
Qy	1639	TTGGTTATGTTGTTAAATTTTCAATGTGAATGCTAAATAGATAGGTGCATATATATTCAAAT	1698
Db	1524	TTGGTTATGTTGTTAAATTTTCAATGTG - AATGTCAAATAGATAGGTGCATATATATTCAAAT	1582
Qy	1699	TTCTTCATTACTTAATGTAATTTGTCATGGCAGGTTTGTTTAAAGTACTATCATGTGTATA	1758
Db	1583	TTCTTCATTACTTAATGTAATTTGTCATGGCAGGTTTGTTTAAAGTACTATCATGTGTATA	1642
Qy	1759	TTTTTGTCAATATTATGTCCCAACAGAAAAATTTCATGTAAAGTCATATTTTTTAAAGGAATAA	1818
Db	1643	TTTTTGTCAATATTATGTCCCAACAGAAAAATTTCATGTAAAGTCATATTTTTTAAAGGAATAA	1702
Qy	1819	ATACATAGCCTTAAAAACAGTGTATACTTTAAATGTAAAAAATAA	1865
Db	1703	ATACATAGCCTTAAAAACAGTGTATACTTTAAATGTAAATGTAACTGACATA	1749

RESULT 4  
AAA30613  
ID AAA30613 standard; cDNA; 1302 BP.

AA  
AC AAA30613;

DT 21-AUG-2000 (first entry)

Human G protein-coupled receptor GPR22 cDNA.

G protein-coupled receptor; GPCR; constitutively active;  
 intracellular loop 3; transmembrane domain 6; drug screening;  
 agonist; antagonist; ss.

OS Homo sapiens.

XX PN WO200022129-A1

XX  
PD  
20-APR-2000.

XX 12-OCT-1999: 99WO-US23938.

PR 13-OCT-1998: 98US-0170496.

PA (AREN-) ARENA PHARM INC.

XX  
PI  
Behan DP, Chalmers DT, Liaw CW

XX  
DR  
WPI: 2000-329165/28

XX  
P-PSDB; AAI90622

PT Non-endogenous constitutively activated human G protein-coupled  
PT receptors, useful for identifying agonists for use as pharmaceutical  
PT agents -

PS Example 1: Page 136-137; 341pp; English.

AA The invention relates to constitutively active, non-endogenous versions  
CC of endogenous human orphan G protein-coupled receptors (GPCRs, AAY90643-  
AA

CC AAY90677 and AAY90683-Y90687), and to DNA encoding them (AAA30709-A30743  
CC mutation in a portion of the protein comprising intracellular loop 3  
CC (IC3) and transmembrane domain 6 (TM6). A non-endogenous amino acid, X,  
CC is substituted for an endogenous residue in IC3 at a position 16 amino  
CC acids N-terminal of an endogenous proline in TM6 to form a sequence  
CC X-(RA)15-Pro. The endogenous amino acid is selected from Lys, His, Arg  
CC or Ala, and is preferably Lys. When the endogenous residue at this  
CC position is Lys, this residue is replaced by His, Arg or preferably Ala.  
CC The 15 amino acid stretch between the substituted amino acid and the Pro  
CC may be endogenous, non-endogenous, or a mixture of endogenous and  
CC non-endogenous residues. The constitutively active GPCRs are useful for  
CC identifying antagonists, agonists and partial agonists for use as  
CC pharmaceutical agents. The mutant proteins are also useful in research  
CC settings for elucidating the roles of the receptors in normal and  
CC diseased conditions. Antagonists for a particular GPCR are useful for  
CC treating diseases and disorders associated with that receptor. Because  
CC the novel mutant GPCRs are constitutively active, they can be used  
CC directly for screening of compounds without the need for endogenous  
CC ligands. The present sequence represents cDNA encoding a human wild-type  
CC GPCR used in an exemplification of the invention. This was cloned and  
CC subjected to site-directed mutagenesis (SDM) to generate DNA encoding  
CC the corresponding mutant of the invention.

XX Sequence 1302 BP; 411 A; 245 C; 226 G; 420 T; 0 other;

Query Match 69.2%; Score 1302; DB 21; Length 1302;  
Best Local Similarity 100.0%; Pred. No. 6.9e-215;  
Matches 1302; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 ATGTGTTTTTCCCATTCGGAATCACATGCGAGTCGATCTGAATCTAAACATTCAGTGGCA 296  
DB 1 ATGTGTTTTTCTCCCATTCGGAATCAACATGCGAGTCGATCTGAATCTAAACATTCAGTGGCA 60  
QY 297 GATGACATTTGATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTT 356  
DB 61 GATGACATTTGATGACATCAACACCAATATGTACCAACCACTATCATATCCGTTAAGCTTTT 120  
QY 357 CAAGTGTCTCTCACCGGATTTCTTATGTTAGAAATTTGTTGGGACTTGGCAGCAACCTC 416  
DB 121 CAAGTGTCTCTCACCGGATTTCTTATGTTAGAAATTTGTTGGGACTTGGCAGCAACCTC 180  
QY 417 ACTGTATTGGTACTTTACTGCGATGAAATCCAACTTAATCAACTCTGTCAGTAACAATATT 476  
DB 181 ACTGTATTGGTACTTTACTGCGATGAAATCCAACTTAATCAACTCTGTCAGTAACAATATT 240  
QY 477 ACAATGAATCTTCATGCTACTTGATGTAAATAATTGTTGGGATGATTCCTCTAACTATA 536  
DB 241 ACAATGAATCTTCATGCTACTTGATGTAAATAATTGTTGGGATGATTCCTCTAACTATA 300  
QY 537 GTTATCCTTCTGCTTTCACTGGAGAGTAACACTGCTCTCATTTGCTGTTTCCATGAGGCT 596  
DB 301 GTTATCCTTCTGCTTTCACTGGAGAGTAACACTGCTCTCATTTGCTGTTTCCATGAGGCT 360  
QY 597 TGTGTATCTTTTGCAAGTGTCTACAGCAATCAACAGTTTTCCTATCTACATTTGGACAGA 656  
DB 361 TGTGTATCTTTTGCAAGTGTCTCAACAGCAATCAACAGTTTTCCTATCTACATTTGGACAGA 420  
QY 657 TATGACATCTCTGTAACAACTTCAACCGAATTCGTACAAATGGCAGAGCTCTAATGTTA 716  
DB 421 TATGACATCTCTGTAACAACTTCAACCGAATTCGTACAAATGGCAGAGCTCTAATGTTA 480  
QY 717 ATGATATCCATTTGGATTTTTTCTTTTTTCTCTTTCTGATTCCTTTTATTCAGGTAAAT 776  
DB 481 ATGATATCCATTTGGATTTTTTCTTTTTTCTCTTTCTGATTCCTTTTATTCAGGTAAAT 540  
QY 777 TTTTTCAGTCTTCAAGTGGAAATACCTGGGAAACAAGACACTTTTATGTGTCAGTACA 836  
DB 541 TTTTTCAGTCTTCAAGTGGAAATACCTGGGAAACAAGACACTTTTATGTGTCAGTACA 600  
QY 837 AATGAATACTACACTGAACCTGGGAATGATATTCACCTCTTGTAGTACAGATCCCAATATTC 896  
DB 601 AATGAATACTACACTGAACCTGGGAATGATATTCACCTCTTGTAGTACAGATCCCAATATTC 660

QY 897 TTTTTCACGTGTTGTAGTAATGTTAATACATACACCAAAATACTTCAAGCTCTTAATATT 956  
DB 661 TTTTTCACGTGTTGTAGTAATGTTAATACATACACCAAAATACTTCAAGCTCTTAATATT 720  
QY 957 CCAATAGGCACAAAGATTTTTCACAGGGCAGAGAGAAAGCAAGAAAGAAAGCAAT 1016  
DB 721 CCAATAGGCACAAAGATTTTTCACAGGGCAGAGAGAAAGCAAGAAAGAAAGCAAT 780  
QY 1017 TCTCTAAACCACACAAATCAGGCTACAGACATGTCCAAAGCAGTGGTGGGAGAAATGTA 1076  
DB 781 TCTCTAAACCACACAAATCAGGCTACAGACATGTCCAAAGCAGTGGTGGGAGAAATGTA 840  
QY 1077 GTCTTTGGTGTGAAGAACTTTCAGTTCTGTGTAATTAATTTGCCCTCCGGGAGCTGTGAAACGA 1136  
DB 841 GTCTTTGGTGTGAAGAACTTTCAGTTCTGTGTAATTAATTTGCCCTCCGGGAGCTGTGAAACGA 900  
QY 1137 CACCGTGAACGACGAGAGAAAGCAAGAAAGAGAGCTTTCAGGATGCTTTTATGATTTTCT 1196  
DB 901 CACCGTGAACGACGAGAGAAAGCAAGAAAGAGAGCTTTCAGGATGCTTTTATGATTTTCT 960  
QY 1197 ACATTTCTTCTCTGCTGGACACCAATTTCTGTTTAAATACCACCAATTTTATGTTTAGGC 1256  
DB 961 ACATTTCTTCTCTGCTGGACACCAATTTCTGTTTAAATACCACCAATTTTATGTTTAGGC 1020  
QY 1257 CCAAGTGACCTTTTAGTAAATTAAGATTTGTTTTTATGTCATGTCCTTATGGAACAAT 1316  
DB 1021 CCAAGTGACCTTTTAGTAAATTAAGATTTGTTTTTATGTCATGTCCTTATGGAACAAT 1080  
QY 1317 ATATTTCACCCCTTATATATATGATTCAGTCTAGACAAAATTTCAAAAGTCTTGAAGAGT 1376  
DB 1081 ATATTTCACCCCTTATATATATGATTCAGTCTAGACAAAATTTCAAAAGTCTTGAAGAGT 1140  
QY 1377 AAAATGAAAAGCGAGTTCTTCTATAGTAGAGCTGATCCCTGCCTAATAATGCTGTA 1436  
DB 1141 AAAATGAAAAGCGAGTTCTTCTATAGTAGAGCTGATCCCTGCCTAATAATGCTGTA 1200  
QY 1437 ATACACAATCTTTGGATAGATCCCAAGAGAAACAAAATTTACCTTTGAAGATAGTGAA 1496  
DB 1201 ATACACAATCTTTGGATAGATCCCAAGAGAAACAAAATTTACCTTTGAAGATAGTGAA 1260  
QY 1497 ATAAGAGAAAAGCTTTAGTGCTCAGGTTGTCACAGACTAG 1538  
DB 1261 ATAAGAGAAAAGCTTTAGTGCTCAGGTTGTCACAGACTAG 1302

RESULT 5  
AAA30722  
ID AAA30722 standard; DNA; 1302 BP.  
XX  
AC AAA30722;  
XX  
DT 21-AUG-2000 (first entry)  
XX  
DE DNA encoding human mutant G protein-coupled receptor GPR22 (F312K).  
XX  
KW G protein-coupled receptor; GPCR; constitutively active;  
KW intracellular loop 3; transmembrane domain 6; drug screening;  
KW agonist; antagonist; mutant; ss.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO200022129-A1.  
XX  
PD 20-APR-2000.  
XX  
PF 12-OCT-1999; 99WO-US23938.  
XX  
PR 13-OCT-1998; 98US-0170496.  
XX  
PA (AREN-) ARENA PHARM INC.  
XX







Db 3 TTTGATTTTCTAGGCAAAACACCACTGCTCCAAAGAAATGTGTTTCTCCTGTTCTGG 62  
 QY 259 AAATCAACATGCGAGTCTGAATCTAACATTAAGTGGAGATGACATTAATGATGACATCAACA 318  
 Db 63 AAATCAACATGCGAGTCTGAATCTAACATTAAGTGGAGATGACATTAATGATGACATCAACA 122  
 QY 319 CCAATATGTAACCAACCACTATCATATCCGTTAAGCTTTTCAAGTGTCTCTCACCGGATTTTC 378  
 Db 123 CCAATATGTAACCAACCACTGTCATACCCCACTTAAGCTTTTCAAGTGTCTCTCACCGGATTTTC 182  
 QY 379 TTATGTTAGAAATGTG 395  
 Db 183 TCATGTTAGAGATCGTG 199

RESULT 8  
 ID ABL33598 standard; DNA; 5487 BP.  
 XX  
 AC ABL33598;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 1571.  
 XX  
 KW Human; Immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; anti-anaemic; cytosine; nontropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.

OS Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 XX 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PF 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-130909/17.  
 XX  
 XX Nucleic acid comprising fragment of chemically modified gene, useful  
 XX for diagnosis and treatment of diseases associated with abnormal  
 XX cytosine methylation -  
 XX  
 PS Claim 1; SEQ ID NO 1571; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 5487 BP; 1532 A; 133 C; 1107 G; 2715 T; 0 other;  
 SQ  
 Query Match 3.6%; Score 67.6; DB 24; Length 5487;  
 Best Local Similarity 47.1%; Pred. No. 0.0079;  
 Matches 208; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

QY 1435 TAATACAACTCTTGGATAGATCCCAAAAGAAACAAAATTTACCTTTCAAGATAGTG 1494  
 Db 780 TAAATACTACTATTAAACATTAACATACTTTCCAAAATTTTCTATACAAATATAT 721  
 QY 1495 AAATAAGAGAAAACGTTTGTAGTCTCAGGTTGTGACAGACTAGAGAAAAGTCTCAGTTT 1554  
 Db 720 AACTTCTTAAACAAATATCCTAATCCATTTATATCTTAAACAAATACCTAAACAA 661  
 QY 1555 CACCAATCCACATTCAAATGAGTTTAAATTTAAATTTGTAACATTAATTTACTGCCA 1614  
 Db 660 AATAATTTATAAATCTATATTATAAATTTATAAATTTATAAATCTATATTATAAAT 601  
 QY 1615 AATATAAGAAAATATTTTAAAGTATGTTTAAATTTTCAATGTCAAATGCTAA 1674  
 Db 600 TATTTATAATTTTATAAATTTTATAAATTTTATAAATTTATAAATTTATAAATTTATAA 541  
 QY 1675 TTAGATAGTCATATATATTAATTTCTTCAATTAATTTTGTGATGTCAGT 1734  
 Db 540 ATATTATATTTAAATTTTAAATTTATATTTTAAATTTAAATTTAAATTTAAATTTAA 481  
 QY 1735 TGTAAAGTACTATCATGTGTATATTTTGAATTTATGTCACAGAGAAAATTTCAATG 1794  
 Db 480 TTTAAATTTAAATTTTATTTAAATTTTAAATTTTATAAATTTATAAATTTATAAATTTATAA 421  
 QY 1795 TAAGTCATATTTTAAAGGATAAATACATAGCCTTAAACAGTGTATACCTTTTAAATG 1854  
 Db 420 TATTTAAATTTTATAAATTTTATAAATTTTATAAATTTTATAAATTTTATAAATTTATAA 361  
 QY 1855 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1876  
 Db 360 TATTAATTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 339

RESULT 9  
 ABO67075/c  
 ID ABO67075 standard; DNA; 7857 BP.  
 XX  
 AC ABO67075;  
 XX  
 DT 28-AUG-2002 (first entry)  
 XX  
 XX Human angiogenesis associated polynucleotide SEQ ID NO 105.  
 DE  
 DE Human; angiogenesis; methylation; eye disease; glaucoma; tumour;  
 KW inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;  
 KW macular degeneration; inflammatory bowel disease; Crohn's disease;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiarteriosclerotic; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200246454-A2.  
 XX  
 PD 13-JUN-2002.  
 XX  
 PF 06-DEC-2001; 2001WO-EPI4320.  
 XX  
 XX 06-DEC-2000; 2000DE-1061338.  
 XX  
 XX (EPIG-) EPIGENOMICS AG.  
 XX  
 XX Schacht O;  
 XX  
 DR WPI; 2002-500450/53.  
 XX  
 PT New nucleic acid fragments from chemically treated  
 PT angiogenesis-associated genes, useful for determining methylation  
 PT status, e.g. in diagnosis or treatment of cancer -  
 XX  
 PS Claim 1; SEQ ID NO 105; 41pp + Sequence Listing; German.  
 XX  
 XX The invention relates to a nucleic acid (I) comprising a segment of 18  
 CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)



Db 1758 AAAAAATCCAAACCAACAAAGACAAATACAAAAAATAAATCTTTCTTCAATCCTA 7099

Qy 1820 TACATAGCCCTTAAACAGCTGTATTAATCTTTAAATGTAAAAAATAAATAAATAAATAA 1878

Db 7098 AAATTAACAATAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7040

RESULT 11

AAS63334/C

ID AAS63334 standard; DNA; 8842 BP.

XX AC AAS63334;

XX DT 29-JAN-2002 (first entry)

XX DE Chemically pretreated metabolism associated gene #29.

XX DE Human; cytostatic; anti-tumour; metabolism; metabolic disease; liver;

KW solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;

KW single nucleotide polymorphism detection; SNP; stool; urine; lung;

KW cerebral-spinal fluid; intestine; brain; heart; prostate; breast;

KW DUSP2; EPHX2; QDPR; SGSH; SHMT2; SLC7A2; SLC7A4; TYMS; ds.

XX OS Homo sapiens.

XX PN WO200176451-A2.

XX PD 18-OCT-2001.

XX PF 06-APR-2001; 2001WO-EP04016.

XX PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2002-010834/01.

XX DR New nucleic acid, useful for diagnosis and therapy of metabolic

PT disease, solid tumour and cancers, comprises segment of chemically

PT modified genomic sequences of genes associated with metabolism -

XX

PS Claim 1; Page 89-91; 143pp; English.

XX The invention relates to a nucleic acid (I) comprising a sequence at

CC least 18 bases of a segment of the chemically pretreated DNA of genes

CC associated with metabolism such as DUSP2 (NM\_004418), EPHX2 (NM\_001979),

CC QDPR (NM\_000320), SGSH (NM\_000199), SHMT2 (NM\_005412), SLC7A2

CC (NM\_003046), SLC7A4 (NM\_004173) and TYMS (NM\_001071) (all

CC undefined). (I) are useful for diagnosis and therapy of metabolic

CC disease, solid tumours and cancers; as primer oligonucleotides for the

CC amplification of DNA sequences, for detecting the cytosine methylation

CC state and/or single nucleotide polymorphisms (SNPs) in a chemically

CC treated DNA of genes associated with metabolism. An array of (I) is

CC useful for ascertaining genetic and/or epigenetic parameters for the

CC diagnosis and/or therapy of existing diseases or the predisposition to

CC specific diseases by analysing cytosine methylations. The method involves

CC chemically treating genomic DNA sample by a solution of bisulphite,

CC hydrogen sulphite or disulphite such that cytosine bases which are

CC unhydrolysed at the 5th-position are converted to uracil or another base

CC which is dissimilar to cytosine in terms of hybridisation behaviour and

CC amplifying fragments of the chemically pretreated genomic DNA. The

CC genomic DNA is from cells or cellular components which contain DNA.

CC sources of DNA comprising, for e.g. cell lines, biopsies, blood, sputum,

CC stool, urine, cerebral-spinal fluid, tissue embedded in paraffin such as

CC tissue from eye, intestine, kidney, brain, heart, prostate, lung, breast

CC or liver, histologic object slides and their combinations. Genetic

CC parameters are mutations, in particular insertions, deletions, point

CC mutations, inversions and polymorphisms of genes associated with

CC metabolism and sequences further required for their regulation.

CC Epigenetic parameters are in particular cytosine methylations and

CC further chemical modifications of DNA bases of genes associated with

CC metabolism. Further epigenetic parameters include for e.g. the

CC acetylation of histones which correlates with DNA methylation.

CC AAS63306-AAS63373 represent chemically pretreated metabolism associated

CC genes, and related primers of the invention.

XX

SQ Sequence 8842 BP; 2294 A; 186 G; 1955 G; 4405 T; 2 other;

Query Match 3.4%; Score 63.8; DB 24; Length 8842;

Best Local Similarity 44.9%; Pred. No. 0.037;

Matches 242; Conservative 0; Mismatches 297; Indels 0; Gaps 0;

Qy 1340 ATTCACTAGACAAAAATTTCAAAAAGGCTTGAAGTAAGTAAAGGAGGAGTGTGTTTC 1399

Db 7578 AATCTTAATAAAAAAATTAATACTATCTACCAAAATATACCCCTAAATATACAAACAA 7519

Qy 1400 TATAGTAGAAGCTGATCCCTGCTAATAATGCTGTATATACACAACCTCTTTGGATAGATCC 1459

Db 7518 TAAACAAAAAACAATAATACGTATATCTTCAACCTCAATAAACCAATTTTAAACAAAC 7459

Qy 1460 CAAAGAAACAAAAAATTTACCTTTGAAGATAGTGAATAAGAGAAAAAGCTTTAGTGCC 1519

Db 7458 AACAAACACAAACAATAACAATACTAAAAAAAACAACATAAATACTAAATAATCAATATT 7399

Qy 1520 TCAGGTGTCACAGACTAGAGAAAGTCTCAGTTTCACCAATCCCAATTTCAATAGATT 1579

Db 7398 TCAATTAACCTTACAAACAAAAAACAATCCACTAATATATCTCAAAAAAATAAAAAAAT 7339

Qy 1580 TTAATTTAATTTGTAATAAATCTGATTTACTGCCAATATATANGAATAATTTTAAGTAT 1639

Db 7338 TAAAAATTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7279

Qy 1640 TGGTTATGTTGTAATTTTCAATGTAATGCTAAATAGATAGTCAATATATATTTCAATT 1699

Db 7278 TAACAATCAACAATAATATATTTCTTAACAATAAATAAATAAATAAATAAATAAATAA 7219

Qy 1700 TCTTCATTAATTAATGTTGTTGTCATGCGAGTCTTTAAAGTACTATCATGTGTATAT 1759

Db 7218 TTATTATATTAAAAATAAATTTCCAAATACAACTACAAAAAATAAATAAATAAATAA 7159

Qy 1760 TTTGTCAATATTATGTCACAGAAATATTCATGTAAGTCATATTTTAAAGCAATAA 1819

Db 7158 AAAAAATCCAAACCAACCAACGACATAACAAAAAATAAATAAATAAATAAATAAATAA 7099

Qy 1820 TACATAGCCCTTAAACAGTGTATTAATCTTTAAAAATGTAAAAAATAAATAAATAA 1878

Db 7098 AAATTAACAATAAACAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7040

RESULT 12

AAS46788/C

ID AAS46788 standard; DNA; 61020 BP.

XX AC AAS46788;

XX DT 18-DEC-2001 (first entry)

XX DE Tumour suppressor gene derived chemically modified sequence #514.

XX DE Human; tumour suppressor gene; oncogene; antitumour; cytostatic;

KW cancer; tumour; CPG dinucleotide; single-nucleotide polymorphism; SNP;

KW cytosine methylation; ds.

XX OS Homo sapiens.

XX PN WO200168912-A2.

XX PD 20-SEP-2001.

XX PF 15-MAR-2001; 2001WO-EP02955.

PR 15-MAR-2000; 2000DE-1013847.  
PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
PI WPI; 2001-602752/68.  
DR  
XX Fragments of chemically modified genes associated with tumour suppressor  
PT genes and oncogenes, useful in designing primers and probes for  
PT analysing diseases associated with cytosine methylation state e.g.  
PT cancer  
XX  
XX Claim 1; SEQ ID No 514; 27pp; English.  
PS  
XX The invention relates to a nucleic acid comprising a sequence of 18  
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
CC bisulphite, of genes associated with tumour suppression and  
CC oncogenes having a sequence taken from 536 (actually 533 since  
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
CC (SS) and sequences complementary to (SS). The nucleic acid may be a  
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
CC form part of a set of probes for detecting the cytosine methylation state  
CC and/or single nucleotide polymorphisms and also to be used in an  
CC array for analysing diseases associated with CpG dinucleotides e.g.  
CC cancers and tumours. The probes can also be used in a method for  
CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
CC and/or therapy of existing diseases or the predisposition to specific  
CC diseases, by analysing cytosine methylations. The parameters may be  
CC compared to another set of genetic and/or epigenetic parameters, the  
CC differences serving as basis for diagnosis and/or prognosis events which  
CC are disadvantageous to patients. The present sequence is one of the  
CC 533 genomic sequences derived from tumour suppressor genes and  
CC oncogenes. Sequences with even numbered Seq ID numbers are the  
CC complementary sequence of the corresponding odd numbered sequence (e.g.  
CC ID 2 and ID1, ID 536 and ID 535, except for those whose partner sequence  
CC is missing).  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 61020 BP; 17884 A; 634 C; 12537 G; 29965 T; 0 other;

Query Match 3.4%; Score 63.4; DB 22; Length 61020;  
Best Local Similarity 46.7%; Pred. No. 0.049;  
Matches 276; Conservative 0; Mismatches 306; Indels 9; Gaps 2;

QY 1294 TAGTCATGCTTATGGAACAACACTATATTCCACCCCTCTATTATATGCATTCACCTAGACAAA 1353  
DB 49874 TAATAATCATCTTTAAAAAACAACAACTAACACCTATATCCCACTACTCCGTAACCGA 49815  
QY 1354 AATTTCAAAGCTGTGAAAAGTAAATGAAAACGAGTGTGTTCTATAGTAGAGCTG 1413  
DB 49814 AATTAATAAATCACTTAAACCTTAAATAATTAATACTACATAATCTATAATTACACTACT 49755  
QY 1414 ATCCCTCGCTTAATATGCTGTTAATACACAACCTCTTGATAGATCCCAAGAACAAAA 1473  
DB 49754 ATACTACAACTAAACAAACAAATAAACCCTTATTCAAAACAAACAAACAAACCAA 49695  
QY 1474 AATATACCTTTGAAGAT---AGTGAATAAGAGAAAAGCTTTAGTGCTCAGGTTGTCA 1530  
DB 49694 ACAACACCAGGAATCATTTTCACTAAACAAACAACTTAAAAATCTAAAAATAAACHTACCTA 49635  
QY 1531 CAGACTAGAGAAAGCTCAGTTTCACCAAAATCCCAAAATCAATCAATGAGTTTAAATTTAAA 1590  
DB 49634 CCAATATACTAAACTTTCAAAATTAATAAACAATAACATAAAAAAAATTAATAAATAAAA 49575  
QY 1591 TTGTAAAAACACTGATATTACTGCCAAATATAGAAAAAATATTTTAAGTATGTTATGTTG 1650

DB 49574 CAAAACAAATCACAAATACACCTATACAAAAAATAAAAAATTTTAAACATCTTTTCCCTTAA 49515  
QY 1651 TAAATTTTCAATGTGAAATGCTAAATAGATAGGTGCATATATATTCATATTCATTTCTTCACTACT 1710  
DB 49514 CATCTCTAATTAATAAACAACAATAATTTAAACACTAAAAAATAATTAA-----AAATTA 49461  
QY 1711 TAATGTATTTGTGCATGGCAGTGTGTTAAAGTACTATCATGTGTATATTTTGTGCAATAT 1770  
DB 49460 CAAAAAATCATTAATCAAAATACCTTTTAAATAAATTTTATTATATATCCATTTTAAATCAATA 49401  
QY 1771 TATGTCCCAACAGAAATATTCATGTAAAGTCATATTTTAAAGCAATAAATACATAGCCCTT 1830  
DB 49400 TTTCTACAAATTTTCTCATTTTAAATTTCTACTCTCTAAATACCATACCTAAAAAT 49341  
QY 1831 AAAACAGTGTATACTTTTAAATGTAAAAAATAAAAAAATAAAAAAATAAAAAA 1881  
DB 49340 ACAACAATATAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 49290

RESULT 13  
ABL32616/c  
ID ABL32616 standard; DNA; 11555 BP.  
XX ABL32616;  
AC ABL32616;  
XX  
DT 26-MAR-2002 (first entry)  
XX Human immune system associated gene SEQ ID NO: 589.  
DE  
XX Human; immune system disease; cytosine methylation; antiasthmatic;  
KW antiarteriosclerotic; antianaemic; cytosine methylation; antiasthmatic;  
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO200200928-A2.  
PN  
XX 03-JAN-2002.  
PD  
XX  
PF 02-JUL-2001; 2001WO-EP07537.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
PI WPI; 2002-130909/17.  
DR  
XX  
PT Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation  
XX  
PS Claim 1; SEQ ID NO 589; 32pp + Sequence Listing; German.  
XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.  
XX  
SQ Sequence 11555 BP; 2890 A; 134 C; 2788 G; 5737 T; 6 other;

```
Query Match      3.3%; Score 63; DB 24; Length 11555;
Best Local Similarity 47.9%; Pred. No. 0.051;
Matches 281; Conservative 0; Mismatches 295; Indels 11; Gaps 3;

QY 1303 CTTATGAACACATATATTTACCCCTCTATATATATGATGACATAGACAAAATTTCAAA 1362
DB 3969 CTCATCCACACAAAAAATTCCTAAATTAATATCATCTATAAATAAATAAATAA 3910
QY 1363 AGGCTCTTGA- AAGTAAATGAAAGGAGGAGTCTTCTATAGTAGAGCTGATCCCTG 1421
DB 3909 AAAACTTTAACTACTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3850
QY 1422 CCTAATATGCTGTAAATACAACTCTTGGATAGATCCCAAGAAACAAAATATACC 1481
DB 3849 ATATCAATATCATCAACACAAACATCTATTTAACCAAAATATAAATAAATAA 3790
QY 1482 TTTGAAGATAGTGAATTAAGAGAAAAACGTTTACTGCTCAGGTTGTCACAGACTAGAGA 1541
DB 3789 ATATTAATAATAATTAATACTACAACAAACATTTTACAAACCAATATATCTTATCACTAA 3730
QY 1542 AAAGTCTCAGTTTCCACCAATCCACATTCACATGAGTTTAAATTTAAATTTCAAAACT 1601
DB 3729 AAAATCCCTAAATAATTAATACTACAACAAACATTTTACAAACCAATATATCTTATCACTAA 3670
QY 1602 GATATTACTGCAA-----ATATAAGAAAAATATTTAAAGTATTTGGTTATGTTGTA 1654
DB 3669 ATATAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3610
QY 1655 TTTTCAATGTGAATGCTAAATAGATAGTGCATATATATTTCAATTTCTTCAATTTAA 1714
DB 3609 TATAATCATCTAAACGAAAAATAAATAAAGCAAAATAAATAAATAAATAAATAAATAA 3550
QY 1715 GTATTTGTTGTCAGTCTTGTAAAGTACTATCATGTGTATATTTTGTCTCATATATG 1774
DB 3549 AACTTAAACTAATAATTTCAATAA---TCTTTAACTCTTAAATAAATAAATAAATAA 3493
QY 1775 TCCACACAGAAAAATATTCATGTAAAGTCATATTTTAAAGGAATAAATAAATAAATAA 1834
DB 3492 ATAATATAAACAACAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3433
QY 1835 CAGTGTATTAATTTTAAATGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 1881
DB 3432 AACTCTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3386

RESULT 14
AAD28380/c
ID AAD28380 standard; DNA; 11555 BP.
XX AC AAD28380;
XX DT 22-APR-2002 (first entry)
XX DE Human chemically treated genomic DNA #21.
XX KW Human; cytostatic; antidepressant; neuroleptic; nootropic; antiaddictive;
XX KW adrenergic alpha-1c-receptor; cytosine methylation; therapy; alcoholism;
XX KW behavioural disorder; neurological; psychiatric; cancer; schizophrenia;
XX KW Tourette's syndrome; smoking; human immunodeficiency virus dementia;
XX KW drug abuse; migraine; ds.
XX OS Homo sapiens.
XX PN WO200202809-A2.
XX PD 10-JAN-2002.
XX PF 02-JUL-2001; 2001WO-EP07540.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PA (EPIG-) EPIGENOMICS AG.
```

```
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-154759/20.
XX Novel nucleic acid useful for diagnosis and therapy of behavioral
XX disorder, neurological disorder and cancer, comprises a sequence of a
XX segment of chemically pretreated DNA of adrenergic alpha-1c-receptor
XX gene.
XX Claim 1; Page 102-108; 190pp; English.
XX The invention relates to nucleic acids comprising a segment of chemically
XX pretreated DNA of adrenergic alpha-1c-receptor gene. The invention also
XX relates to oligonucleotides or peptide nucleic acid (PNA) oligomers
XX useful for detecting cytosine methylations. The pretreated DNA is useful
XX for the diagnosis or therapy of behavioural disorders, neurological
XX disorders and cancer, in particular major depressive disorder, Tourette's
XX syndrome, schizophrenia, psychiatric and neurological disorders, smoking,
XX drug abuse, alcoholism, personality traits, compulsive gambling, human
XX immunodeficiency virus dementia, migraine, behaviours in schizophrenia
XX and schizoaffective patients, and suicidal behaviour in patients with
XX schizophrenia. The nucleic acid is useful for detecting the methylation
XX state of all CpG dinucleotides and/or single nucleotide polymorphisms
XX (SNPs). The present sequence is human chemically treated genomic DNA.
XX SQ Sequence 11555 BP; 2890 A; 134 C; 2788 G; 5737 T; 6 other;

Query Match      3.3%; Score 63; DB 24; Length 11555;
Best Local Similarity 47.9%; Pred. No. 0.051;
Matches 281; Conservative 0; Mismatches 295; Indels 11; Gaps 3;

QY 1303 CTTATGAACACATATATTTACCCCTCTATATATATGATGACATAGACAAAATTTCAAA 1362
DB 3969 CTCATCCACACAAAAAATTCCTAAATTAATATCATCTATAAATAAATAAATAAATAA 3910
QY 1363 AGGCTCTTGA- AAGTAAATGAAAGGAGGAGTCTTCTATAGTAGAGCTGATCCCTG 1421
DB 3909 AAAACTTTAACTACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3850
QY 1422 CCTAATATGCTGTAAATACAACTCTTGGATAGATCCCAAGAAACAAAATATACC 1481
DB 3849 ATATCAATATCATCAACACAAACATCTTATTTAACCAAAATATAAATAAATAAATAA 3790
QY 1482 TTTGAAGATAGTGAATTAAGAGAAAAACGTTTACTGCTCAGGTTGTCACAGACTAGAGA 1541
DB 3789 ATATTAATAATAATTAATACTACAACAAACATTTTACAAACCAATATATCTTATCACTAA 3730
QY 1542 AAAGTCTCAGTTTCCACCAATCCACATTCACATGAGTTTAAATTTAAATTTCAAAACT 1601
DB 3729 AAAATCCCTAAATAATTAATACTACAACAAACATTTTACAAACCAATATATCTTATCACTAA 3670
QY 1602 GATATTACTGCAA-----ATATAAGAAAAATATTTAAAGTATTTGGTTATGTTGTA 1654
DB 3669 ATATAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3610
QY 1655 TTTTCAATGTGAATGCTAAATAGATAGTGCATATATATTTCAATTTCTTCAATTTAA 1714
DB 3609 TATAATCATCTAAACGAAAAATAAATAAAGCAAAATAAATAAATAAATAAATAAATAA 3550
QY 1715 GTATTTGTTGTCAGTCTTGTAAAGTACTATCATGTGTATATTTTGTCTCATATATG 1774
DB 3549 AACTTAAACTAATAATTTCAATAA---TCTTTAACTCTTAAATAAATAAATAAATAA 3493
QY 1775 TCCACACAGAAAAATATTCATGTAAAGTCATATTTTAAAGGAATAAATAAATAAATAA 1834
DB 3492 ATAATATAAACAACAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3433
QY 1835 CAGTGTATTAATTTTAAATGTAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 1881
DB 3432 AACTCTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3386
```







2:

Qy	139	TCTTACAAAGAACACGTTTACGTCAATTTAAATTCGCAAAATATCAAATPAGTTATTCTAT	198
Db	28	TATTTGGAGAAAACGAAGAATCTCTCAGACACATGAAATAAACAACACAGCTACTTC	- 84
Qy	199	TTCACITTTCTAGGGAAAAAACCACACTGCTCCAAAGAATGTGTTTTTCTCCCATTTCTGG	258
Db	85	-ACTGTTGTTCAGGNAAAAAACCACTGCTCCAAAGAATGTGTTTTTCTCCCATTTCTGG	143
Qy	259	AATTAACATGCAGTCTGAATCTAACTTTACAGTGCAGAGTGACATTTGATGACATCAACA	318
Db	144	AAATCAACATGCAGTCTGAATCTAACTTTACAGTGCAGAGTGACATTTGATGACATCAACA	203
Qy	319	CCAATATGTACCAACCACTATCATATCGTTAAAGCTTTCAAGTGTCTCTCACCGGATTTTC	378
Db	204	CCAATATGTACCAACCACTATCATATCGTTAAAGCTTTCAAGTGTCTCTCACCGGATTTTC	263
Qy	379	TTATGTTAGAAATGTGTTGGGACTTTGGCAGCAACCTCACTGTATTTGGTAGTTTACTGCA	438
Db	264	TTATGTTAGAAATTTTTTTTGGACTTTGGCAGCAACCTCACTGTATTTGGTAGTTTACTGCA	323
Qy	439	TGAAATCCAACTTAATCAACTCTGTCACTTAACATTTACAAATGAATCTTCATCTACTTG	498
Db	324	TGAAATCCAACTTAATCAACTCTGTCACTTAACATTTACAAATGAATCTTCATCTACTTG	383
Qy	499	ATGTAATAATTTGTGGGATGTATTCCTCACTAACTATAGTTATGCTTCTGCTTTCACCTGG	558
Db	384	ATGTAATAATTTGTGGGATGTATTCCTCACTAACTATAGTTATGCTTCTGCTTTCACCTGG	443
Qy	559	AGAGTAACACTGCTCTCAATTTGCTGTTTCCATGAGGCTTGTATCTTTTGTGCAAGTGTCT	618
Db	444	AGAGTAACACTGCTCTCAATTTGCTGTTTCCATGAGGCTTGTATCTTTTGTGCAAGTGTCT	503
Qy	619	CAACAGCAATCAACGTTTTTGTCTATCACTTTTGGCAGAGATGACATCTCTGTAAACCTG	678
Db	504	CAACAGCAATCAACGTTTTTGTCTATCACTTTTGGCAGAGATGACATCTCTGTAAACCTG	563
Qy	679	CAAACCGAATTTCTGACAATGGGCAGAGCTGTAATGTTAATGATATCCATTTGGATTTTTT	738
Db	564	CAAACCGAATTTCTGACAATGGGCAGAGCTGTAATGTTAATGATATCCATTTGGATTTTTT	623
Qy	739	CTTTTTTCTCTTCTCGATTCCTTTTATTTGAGGTAAATTTTTTTCAGTCTTCAAAGTGGAA	798
Db	624	CTTTTTTCTCTTCTCGATTCCTTTTATTTGAGGTAAATTTTTTTCAGTCTTCAAAGTGGAA	683
Qy	799	ATACCTGGGNAACAGACACACTTTTATGTGTCAGTACAAATGAATACTACACTGAACCTGG	858
Db	684	ATACCTGGGNAACAGACACACTTTTATGTGTCAGTACAAATGAATACTACACTGAACCTGG	743
Qy	859	GAAATGATTTATCACCTGTTAGTACAGATCCCCAATATCTTTTTTCACTGTTGTAGTAATGT	918
Db	744	GAAATGATTTATCACCTGTTAGTACAGATCCCCAATATCTTTTTTCACTGTTGTAGTAATGT	803
Qy	919	TAATCATATACACCAAAATCTTCAGGCTCTTAATATTTTCGAATAGGCACAGATTTTCAA	978
Db	804	TAATCATATACACCAAAATCTTCAGGCTCTTAATATTTTCGAATAGGCACAGATTTTCAA	863
Qy	979	CAGGGCAGAGAGAAAGCAAGAAAGACAAATTTCTCTTAACACACACAATCAGG	1038
Db	864	CAGGGCAGAGAGAAAGCAAGAAAGACAAATTTCTTAACACACACAATCAGG	923
Qy	1039	CTACAGACATGTCACAAACGACTGTTGGGAGAAATGTAGTCTTTGGTGTAGAACTTCAG	1098
Db	924	CTACAGACATGTCACAAACGACTGTTGGGAGAAATGTAGTCTTTGGTGTAGAACTTCAG	983
Qy	1099	TTTCTGTATAATTTGCCCTCCGCGGAGCTGTGAAACGACACCGCTGAACGACGAGAAAGAC	1158
Db	984	TTTCTGTATAATTTGCCCTCCGCGGAGCTGTGAAACGACACCGCTGAACGACGAGAAAGAC	1043
Qy	1159	AAAAGAGAGTCTTCAGGATGTCTTTATTTGATTTATTTCTTCTCTCTCTCTCTCTCTCTCT	1218
Db	1044	TAAAGAGAGTCTTCAGGATGTCTTTATTTGATTTATTTCTTCTCTCTCTCTCTCTCTCTCT	1103

Qy	1219	CAATTCCTGTTTAAATACACACATTTTATGTTTAGGCCCAAGTGACCTTTTAGTAAAT	1278
Db	1104	CAATTTCTGTTTAAATACACCATTTTATGTTTAGGCCCAAGTGACCTTTTAGTAAAT	1163
Qy	1279	TAAGATGTGTTTATTTAGTCATGGCTTATGGAACAACATATATTCACCCCTCTATTATATG	1338
Db	1164	TAAGATGTGTTTATTTAGTCATGGCTTATGGAACAACATATATTCACCCCTCTATTATATG	1223
Qy	1339	CAATCTAGACAAAAATTTCAAAGGCTCTTGAAAGTAAATGAAAAAGCGAGCTGTGT	1398
Db	1224	CAATCTAGACAAAAATTTCAAAGGCTCTTGAAAGTAAATGAAAAAGCGAGCTGTGT	1283
Qy	1399	CTATAGTAGAAGTGATCCCGTCCTTAATAATGCTGTGAATACACAACCTTCCTGGATAGATC	1458
Db	1284	CTATAGTAGAAGTGATCCCGTCCTTAATAATGCTGTGAATACACAACCTTCCTGGATAGATC	1343
Qy	1459	CCAAAAGAAAACAAAAATTTACTTTTGAAGTAGTGAATAGAGAAAAGCGTTTACTGTC	1518
Db	1344	CTAAAAGAAAACAAAAATTTACTTTTGAAGTAGTGAATAGAGAAAAGCTTTTAGTGC	1403
Qy	1519	CTCAGGTTGTACAGACTAGAGAAAAGTCTCAGTTTCCACCAATCCACATTCAAATGAGT	1578
Db	1404	CTCAGGTTGTACAGACTAGAGAAAAGTCTCAGTTTCCACCAATCCACATTCAAATGAGT	1463
Qy	1579	TTTAAATTTAAATGTAAAAACGTATATTACTGTCACCAATATAAGAAAAATATTTTAGTA	1638
Db	1464	TTTAAATTTAAATGTAAAAACGTATATTACTGTCACCAATATAAGAAAAATATTTTAGTA	1523
Qy	1639	TTGGTTATGTTGTAATTTTCAATGTGAATGCTCAATTAGATAGGTGCATATATATTCAT	1698
Db	1524	TTGGTTATGTTGTAATTTTCAATGTG - AATGTCATTTAGATAGGTGCATATATATTCAT	1582
Qy	1699	TTCTTCATTACTTAATGTATTTGTTCGATGGCAGTTTGTAAAGTACATCATGTGTATA	1758
Db	1583	TTCTTCATTACTTAATGTATTTGTTCGATGGCAGTTTGTAAAGTACATCATGTGTATA	1642
Qy	1759	TTTTTGTCATTAATATGTCACAAGAAAAATATTCATGTAAAGTCATATTTTTTAAAGSAATA	1818
Db	1643	TTTTTGTCATTAATATGTCACAAGAAAAATATTCATGTAAAGTCATATTTTTTAAAGSAATA	1702
Qy	1819	ATACATAGCCCTTAAAAACAGTGTATACTTTAAAAATGTAAAAAAAATA 1865	
Db	1703	ATACATAGCCCTTAAAAACAGTGTATACTTTAAAAATGTAAATGTAACTGACATA 1749	

RESULT 2  
 US-07-867-106-2  
 ; Sequence 2, Application US/07867106  
 ; Patent No. 5389526  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Slade, Martin B  
 ; APPLICANT: Chang, Andy C M  
 ; APPLICANT: Williams, Keith L  
 ; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular  
 ; TITLE OF INVENTION: Slime Moulds of the Genus Dictyostellium  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 53895  
 ; STREET: One Liberty Place 46th Floor  
 ; CITY: Philadelphia  
 ; STATE: PA  
 ; COUNTRY: USA  
 ; ZIP: 19103  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/07/867,106  
 ; FILING DATE: 19920625  
 ; PRIOR APPLICATION DATA:

[illegible]

Db 5599 TAAA 5596

## RESULT 4

```

US-08-487-826B-13
: Sequence 13, Application US/08487826B
: Patent No. 5993827
: GENERAL INFORMATION:
: APPLICANT: Sim, Kim L.
: APPLICANT: Chitnis, Chetan
: APPLICANT: Miller, Louis H.
: APPLICANT: Peterson, David S.
: APPLICANT: Su, Xin-zhaun
: APPLICANT: Wellens, Thomas E.
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe Martens Olson & Bear
: STREET: 620 Newport Center Drive 16th Floor
: CITY: Newport Beach
: STATE: California
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/487,826B
: FILING DATE: 10-SEP-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Israelson, Ned
: REGISTRATION NUMBER: 29,655
: REFERENCE/DOCKET NUMBER: NIH121.001CPI
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 19124 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
US-08-487-826B-13

```

Query Match	2.98;	Score 54.4;	DB 2;	Length 19124;
Best Local Similarity	46.7%;	Pred. No. 0.019;		
Matches 209;	Conservative	0;	Mismatches 236;	Indels 3; Gaps 1;
Qy 1437	ATACACAACTCTTGGATAGATCCCAAGAAACAAAAAAATTCCTTTGAAGATAGTCAA	1496		
Db 15472	ATAAAAAAAATTTATTAATAATAAAAAAAAGGAGAAAAATTTTTTAA	15531		
Qy 1497	ATAAGAGAAAAACGTTTAGTGCCCTCAGGTTGTCCACAGACTAGAGAAAAAGTCTCAGTTTCA	1556		
Db 15532	AAAAATAAAAAATTATAATAAATAATAAAATTTGTATAGATAAAAAATGAAGAAGATT	15591		
Qy 1557	CCAAATCCACATCCAAATGAGCTTTTAAATTTAAATTTGTAACACTGATATTACTGCGCCAAA	1616		
Db 15592	TCAAAAAAAATTTAAAAAAAATTTTATATAAAAAAAATGATTATAAAAAAAATAAAA	15651		
Qy 1617	TATAAGAAAAATATTTTAAAGTATTTGGTTATGTTGTAATTTTCAATGGAATGCTAATT	1676		
Db 15652	ACAAAGAGAAAAAAAACATTAATAAAAAAAATATATATCATATAAAAAACAAAAA	15711		
Qy 1677	AGATAGGTCATATATATTCATTTTC---TTCATTACTTAATGTATTTGTTGCATGGCAGT	1733		

Db	15712	ANAAGAAAAAATATATTAAATAAAAAATATATATCATAAATAAAAAAATAAAAAA	15771
Qy	1734	TTGTTAAAGTACTATCATGCTGTATATTTTGGCAATATTATGTCACACAGAAAAATATTCAT	1793
Db	15772	ATGTTTAAAAAATAATATATCATATAATAAAAAAATTTATTTAATAAAAAAATAATAA	15831
Qy	1794	GTAAGTCATATTTTTTAAGGGAATAAATACATAGCGCTTAAAAACAGCTGTATAACTTTTAAAAAT	1853
Db	15832	TAAATAAAAAATTTAATTAATAAAAAAATAATAATAAAAAAATAATTAAATTAATAATAA	15891
Qy	1854	GTAAAAAATAAAAAAATAAAAAAATAAAAAA	1881
Db	15892	AAAAAAATTAATAAAAAATTTAATGGAATA	15919

RESULT 5

```

US-09-182-816-22
; Sequence 22, Application US/09182816
; Patent No. 6143542
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-C1
; CURRENT APPLICATION NUMBER: US/09/182,816
; CURRENT FILING DATE: 1998-10-29
; EARLIER APPLICATION NUMBER: 08/989,510
; EARLIER FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 22
; LENGTH: 1736
; TYPE: DNA
; ORGANISM: Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (159)..(1553)
US-09-182-816-22

```

Query Match	2.88;	Score 53;	DB 3;	Length 1736;
Best Local Similarity	59.7%;	Pred. No. 0.024;		
Matches 89;	Conservative 0;	Mismatches 60;	Indels 0;	Gaps 0;
QY 1733	TTTGTGTTAAAGTACATCATGTGTATATTTTGTGCAATATATTATGTCCAACAGAAAAATATTCA	1792		
Db 1574	TTTGTGTATATATATATATATGTATTAATAATAATGTAATTACTGTGAATATAAACGATATGGA	1633		
QY 1793	TGTAAGTCATATTTTAAAGGAATAAATACATAGCTTTAAACACAGTGATTAACTTTAAAA	1852		
Db 1634	TTTTTATTTCAAACTTGTGCAATATATAAAAAAATAAAAAAATAAAAAAATAAAAAA	1693		
QY 1853	TGTAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	1881		
Db 1694	AAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA	1722		

RESULT 6

```

RESOLUT. O
US-09-182-816-24/c
; Sequence 24, Application US/09182816
; Patent No. 6143542
; GENERAL INFORMATION:
; APPLICANT: Wisniewski, Nancy
; APPLICANT: Silver, Gary M.
; APPLICANT: Lo, Katherine C.
; APPLICANT: Brandt, Kevin S.
; TITLE OF INVENTION: NOVEL FLEA EPOXIDE HYDROLASE NUCLEIC ACID MOLECULES,
; TITLE OF INVENTION: PROTEINS AND USES THEREOF
; FILE REFERENCE: FC-3-CI
; CURRENT APPLICATION NUMBER: US/09/182,816
; CURRENT FILING DATE: 1998-10-29

```

; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
US-09-182-816-24

Query Match 2.8%; Score 53; DB 3; Length 1736;  
Best Local Similarity 59.7%; Pred. No. 0.024;  
Matches 89; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
QY 1733 TTTGTTAAAGTACTATCATGCTGATATTTTGTCAATATTATGTCCACAGAAATATTTCA 1792  
DB 163 TTTGTGATAATAATATATGTTAAATAATAATGTAATTACTGTGAATAAACGATATGGA 104  
QY 1793 TGTAAAGTCATATTTTAAAGGAATAATACATAGCCCTTAAACAGTGTATTAACCTTTAAAA 1852  
DB 103 TTTTATTCAACTGTGCAATATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 44  
QY 1853 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1881  
DB 43 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 15

## RESULT 7

US-09-471-528-22  
; Sequence 22, Application US/09471528  
; Patent No. 6153397  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1-1  
; CURRENT APPLICATION NUMBER: US/09/471,528  
; CURRENT FILING DATE: 1999-12-27  
; EARLIER APPLICATION NUMBER: 09/182,816  
; EARLIER FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (159)..(1553)  
US-09-471-528-22

Query Match 2.8%; Score 53; DB 3; Length 1736;  
Best Local Similarity 59.7%; Pred. No. 0.024;  
Matches 89; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
QY 1733 TTTGTTAAAGTACTATCATGCTGATATTTTGTCAATATTATGTCCACAGAAATATTTCA 1792  
DB 1574 TTTGTGATAATAATAATGTTAAATAATAATGTAATTACTGTGAATAAACGATATGGA 1633  
QY 1793 TGTAAAGTCATATTTTAAAGGAATAATACATAGCCCTTAAACAGTGTATTAACCTTTAAAA 1852  
DB 1634 TTTTATTCAACTGTGCAATATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1693  
QY 1853 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1881  
DB 1694 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1722

## RESULT 8

Query Match 2.8%; Score 53; DB 4; Length 1736;  
Best Local Similarity 59.7%; Pred. No. 0.024;

US-09-471-528-24/C  
; Sequence 24, Application US/09471528  
; Patent No. 6153397  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1-1  
; CURRENT APPLICATION NUMBER: US/09/471,528  
; CURRENT FILING DATE: 1999-12-27  
; EARLIER APPLICATION NUMBER: 09/182,816  
; EARLIER FILING DATE: 1998-10-29  
; EARLIER APPLICATION NUMBER: 08/989,510  
; EARLIER FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 24  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
US-09-471-528-24

Query Match 2.8%; Score 53; DB 3; Length 1736;  
Best Local Similarity 59.7%; Pred. No. 0.024;  
Matches 89; Conservative 0; Mismatches 60; Indels 0; Gaps 0;  
QY 1733 TTTGTTAAAGTACTATCATGCTGATATTTTGTCAATATTATGTCCACAGAAATATTTCA 1792  
DB 163 TTTGTGATAATAATAATGTTAAATAATAATGTAATTACTGTGAATAAACGATATGGA 104  
QY 1793 TGTAAAGTCATATTTTAAAGGAATAATACATAGCCCTTAAACAGTGTATTAACCTTTAAAA 1852  
DB 103 TTTTATTCAACTGTGCAATATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 44  
QY 1853 TGTAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1881  
DB 43 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 15

## RESULT 9

US-09-634-530-22  
; Sequence 22, Application US/09634530  
; Patent No. 6290958  
; GENERAL INFORMATION:  
; APPLICANT: Wisniewski, Nancy  
; APPLICANT: Silver, Gary M.  
; APPLICANT: Lo, Katherine C.  
; APPLICANT: Brandt, Kevin S.  
; TITLE OF INVENTION: FLEA EPOXIDE HYDROLASE PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-3-C1-1  
; CURRENT APPLICATION NUMBER: US/09/634,530  
; CURRENT FILING DATE: 2000-08-08  
; PRIOR APPLICATION NUMBER: 09/471,528  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: 09/182,816  
; PRIOR FILING DATE: 1998-10-29  
; PRIOR APPLICATION NUMBER: 08/989,510  
; PRIOR FILING DATE: 1997-12-12  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 22  
; LENGTH: 1736  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (159)..(1553)  
US-09-634-530-22

Query Match 2.8%; Score 53; DB 4; Length 1736;  
Best Local Similarity 59.7%; Pred. No. 0.024;





NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRADEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pT9gpt-F1s  
US-08-232-463-14

Query Match 2.7%; Score 51; DB 1; Length 7218;  
Best Local Similarity 12.7%; Pred. No. 0.078;  
Matches 48; Conservative 167; Mismatches 162; Indels 0; Gaps 0;  
QY 333 CCACATATCATATCCGTTTCAAGTCTCTCACCGGATTTCTTATGTAGAAATT 392  
Db 1149 YY 1208  
QY 393 GTGTTGGGACTTGGCAGCACTCTACTGTATTGGTACTTACTGATGAAATCAACTTA 452  
Db 1209 YY 1268  
QY 453 ATCAACTCTGTCTCAGTAACATTTATCAATGAATCTTCAATGATGATTAATTTGT 512  
Db 1269 YY 1328  
QY 513 GTGGGATGATTCCTTAAGTATAGTTATCTCTTCTACTGAGAGTAACACTGCT 572  
Db 1329 YY 1388  
QY 573 CACATTTGCTGTTCCATCAGGCTTGTGTATCTTTTGGCAAGTGTCTCACACATCAAC 632  
Db 1389 YY 1448  
QY 633 GTTTTGTCTATCATTGTCAGATATGACATCTCTGTAACACCTGCAACCGAATTCG 692  
Db 1449 TATCTCTTAAGTCTGATGATAGTAAATACAGTATGCTTACATGCGGTTTGTG 1508  
QY 693 ACAATGGGACAGCTGT 709  
Db 1509 AAACGAATAGATGCGT 1525

RESULT 13  
PCT-US95-00362-1  
; Sequence 1, Application PC/TUS9500362  
; GENERAL INFORMATION:  
; APPLICANT: IMMUNEX CORPORATION  
; TITLE OF INVENTION: Ligand That Binds Fas Antigen  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: US  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple 7.1  
; SOFTWARE: Microsoft Word, Version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/00362  
; FILING DATE: 06-JAN-1995  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/179,138  
; FILING DATE: 07-JAN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/190,559  
; FILING DATE: 01-FEB-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Anderson, Kathryn A.  
; REGISTRATION NUMBER: 32,172  
; REFERENCE/DOCKET NUMBER: 2805-WO  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1841 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 93..938  
PCT-US95-00362-1

Query Match 2.7%; Score 50; DB 5; Length 1841;  
Best Local Similarity 53.9%; Pred. No. 0.099;  
Matches 125; Conservative 0; Mismatches 105; Indels 2; Gaps 1;  
QY 1652 AAATTTCAATGTAATGCTAATTAGATAGGTGATATATATATATATATTTCTTCAATTT 1711  
Db 1592 ATATTGTGAAGTACATATTAGGAAATATGGCTTGCATTTGCTCAAGATTTTGAATGCTT 1651  
QY 1712 AATGATTTGTTGTCAGGAGTTTGTAAAGTACT--ATCATGTGATATATTTTGTCAATA 1769  
Db 1652 CCGTCAACTCAACTCAATTAATGCTTAAATAATCATTTGATGCTAGTACTAATGATTTT 1711  
QY 1770 TTATGTCCCAACAGAAATATTCATGTAAGTCATATTTTAAAGGAATAATACATAGCCT 1829  
Db 1712 TCCATATATATTAATTAATTTATGTAGATGTGCAATTTTGTGAATGAAACATGTAAT 1771  
QY 1830 TAAACAGTGTATACTTTAAATGTAATAAAAAAAAAAAAAAAAAAAAAA 1881  
Db 1772 AAAAGATATATCTAGGATACAAATAAAAAAAAAAAAAAAAAAAAAA 1823

RESULT 14  
US-08-451-405A-2  
; Sequence 2, Application US/08451405A  
; Patent No. 5736358  
; GENERAL INFORMATION:  
; APPLICANT: FASEL, NICOLAS JOSEPH  
; REYMOND, CHRISTOPHE DOMINIQUE



;; TITLE OF INVENTION: DICTYOSTELID EXPRESSION VECTOR AND  
;; TITLE OF INVENTION: METHOD FOR EXPRESSING A DESIRED PROTEIN  
;; NUMBER OF SEQUENCES: 3  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: THE WEBB LAW FIRM  
;; STREET: 700 KOPPERS BUILDING, 436 SEVENTH AVENUE  
;; CITY: PITTSBURGH  
;; STATE: PENNSYLVANIA  
;; COUNTRY: UNITED STATES OF AMERICA  
;; ZIP: 15219-1818  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: 3.5" FLOPPY DISK  
;; COMPUTER: Midwest Micro 486-50  
;; OPERATING SYSTEM: DOS  
;; SOFTWARE: WORDPERFECT 6.1  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/451,405A  
;; FILING DATE: 26-MAY-1995  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 07/965,273  
;; FILING DATE: 15-JAN-1993  
;; INFORMATION FOR SEQ ID NO: 2:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 731  
;; TYPE: NUCLEIC ACID  
;; STRANDEDNESS: SINGLE  
;; TOPOLOGY: UNKNOWN  
US-08-451-405A-2

Query Match 2.6%; Score 49.8; DB 1; Length 731;  
Best Local Similarity 45.4%; Pred. No. 0.092;  
Matches 303; Conservative 0; Mismatches 352; Indels 12; Gaps 3;  
QY 1220 AATTCTGTTTTAAATACCACTTTATGTTAGGCCCAAGTACCTTTTAGTAAAT 1279  
DB 58 AGTATATGTTTTTTCCTTTTTTTTATTTATTTTATTTTAAATAATAAAT 117  
QY 1280 AGATTGTTTTTTAGTCATGCGCTTATGGAACAATATTTTACCCTCTATTATGC 1339  
DB 118 TAGAATAAATAATTTCTATTGAAGGAGTTTTTTATTTGTTATTTAAATATATATAACAT 177  
QY 1340 ATTCA--CTAGACAAAATTTCAAAGGCTTCGAAAGTAAAA--TGAAAAAGCGAGTT 1394  
DB 178 AGTGAACCTAATAATAGATTGTGACGGTATATGTAAGAAATTTCTAAAAAATAATTC 237  
QY 1395 GTTCTATAGTAGAAGCTGATCCCTGCTTAATGCTGTGTAATACACAACTCTTGATA 1454  
DB 238 AGATAATTTTGGATTGGAACAACAACCAACCAAAAAAATAAAAAAATAAATC 297  
QY 1455 GATCCCAAGAACAACAATAAATTTACCTTTGAAGTAGTGAATAAGAGAAACGTTTA 1514  
DB 298 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 357  
QY 1515 GTCCTCAGTTGTCACAGACTAGAGAAAGTCTCAGTTTCCACAAATCCACATCAAAAT 1574  
DB 358 ATATTATATATATATTTAATTTGCGAAACACACTTTTAAACACACTCTATTATCTACA 417  
QY 1575 GAGTTTTTAAATTTGTAATAAAGCTGATTTACTGCCAAATATATAGAAAAATATTTTA 1634  
DB 418 AAGGTTTTAAATTTTAAATTTTATTTTAAATTTATTTTAAATAAATAAATTTT 477  
QY 1635 AGTATGGTATGTTGTAATTTTCAATGTGAATGCTTAATTAGATAGGTGCATATATATT 1594  
DB 478 ATTTTAAATTTTATTTTATTTTATATATATCAATTTAGAAATAAATAATTTTCCCTATAG 530  
QY 1695 CAATTTCTTCATTACTTAATGTTATTTGTCATGGCAGTTTGTGTAAGTACTATCATGTG 1754  
DB 531 ACAATTTATTTATTTATTTTATATATATCAATTTAGAAATAAATAATTTTCCCTATAG 590  
QY 1755 TATATTTGTCATATATATGTCACAGAAAAATTTATCATGTAAGTCATATTTTAAAGGA 1814  
DB 591 TAGTAATAATAATTTCTTTTTTAATAAATAAATACCTTTTCTACATTTATTTTATTTA 650

QY 1815 ATAAATACATAGCCTTAAACACAGTGTATAACTTTAAATCTAAAAATCTAAAAA 1874  
DB 651 TTTTTCCTTTTAATCATCTCAAAATTTTATTTTAAAAAATAAATAAATAAATAA 710  
QY 1875 AAAAAA 1881  
DB 711 ACAACA 717  
RESULT 15  
US-09-019-942-2  
; Sequence 2, Application US/09019942  
; Patent No. 6033855  
; GENERAL INFORMATION:  
; APPLICANT: Berlin, John  
; TITLE OF INVENTION: GENES ENCODING CASPASE RECRUITMENT  
; TITLE OF INVENTION: DOMAIN POLYPEPTIDES  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/019,942  
; FILING DATE: 06-FEB-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Melkielejohn, Ph.D., Anita L.  
; REGISTRATION NUMBER: 35,283  
; REFERENCE/DOCKET NUMBER: 07334/068001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1931 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
US-09-019-942-2

Query Match 2.6%; Score 49.2; DB 3; Length 1931;  
Best Local Similarity 65.5%; Pred. No. 0.14; Mismatches 0; Gaps 0;  
Matches 72; Conservative 0; Mismatches 38; Indels 0; Gaps 0;  
QY 1772 ATGTCCACAGAAAAATTCATGTAGTCAATATTTTAAAGGAATAATACATAGCCTTA 1831  
DB 1811 ATTACTTCAATAAATAAAGCATGTAGTCAATGTGTTCAGGAAGAAATGTGTTCATAA 1870  
QY 1832 AACAGTGTATACCTTTAAATGTAAAAAATAAATAAATAAATAAATAAATAA 1881  
DB 1871 AAGGATATTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1920

Search completed: December 10, 2002, 09:46:51  
Job time : 161 secs

**This Page Blank (uspto)**